PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/31, C07K 14/35, G01N 33/50, 33/53, 33/58, 33/68, C12Q 1/68, C12N 1/19, 1/21, 5/10, C07K 16/12, 19/00 // (C12N 1/21, C12R 1:19)

A2

(11) International Publication Number: **WO 98/53076**

(43) International Publication Date: 26 November 1998 (26.11.98)

(21) International Application Number: PCT/US98/10514

(22) International Filing Date: 20 May 1998 (20.05.98)

(30) Priority Data:

08/858,998	20 May 1997 (20.05.97)	US
09/073,009	5 May 1998 (05.05.98)	US

(71) Applicant: CORIXA CORPORATION [US/US]; Suite 200, 1124 Columbia Street, Seattle, WA 98104 (US).

(72) Inventors: ALDERSON, Mark, R.; 1116 Grow Avenue Northwest, Bainbridge Island, WA 98110 (US). DILLON, Davin, C.; 21607 Northeast 24th Street, Redmond, WA 98053 (US). SKEIKY, Yasir, A., W.; 8327 25th Avenue N.W., Seattle, WA 98117 (US). CAMPOS-NETO, Antonio; 9308 N.E. Midship Court, Bainbridge Island, WA 98110 (US).

(74) Agent: KOHLER, Thomas, D.; Perkins & Edmunds LLP, 1155 Avenue of the Americas, New York, NY 10036-2711 (US).

(81) Designated States: AL, AM, AT, AU, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: COMPOUNDS FOR DIAGNOSIS OF TUBERCULOSIS AND METHODS FOR THEIR USE

(57) Abstract

Compounds and methods for diagnosing tuberculosis are disclosed. The compounds provided include polypeptides that contain at least one antigenic portion of one or more *M. tuberculosis* proteins, and DNA sequences encoding such polypeptides. Diagnostic kits containing such polypeptides or DNA sequences and a suitable detection reagent may be used for the detection of *M. tuberculosis* infection in patients and biological samples. Antibodies directed against such polypeptides are also provided.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

Description

COMPOUNDS FOR DIAGNOSIS OF TUBERCULOSIS AND METHODS FOR THEIR USE

Technical Field

The present invention relates generally to the detection of *Mycobacterium tuberculosis* infection. The invention is more particularly related to polypeptides comprising a *Mycobacterium tuberculosis* antigen, or a portion or other variant thereof, and the use of such polypeptides for the serodiagnosis of *Mycobacterium tuberculosis* infection.

Background of the Invention

Tuberculosis is a chronic, infectious disease, that is generally caused by infection with *Mycobacterium tuberculosis*. It is a major disease in developing countries, as well as an increasing problem in developed areas of the world, with about 8 million new cases and 3 million deaths each year. Although the infection may be asymptomatic for a considerable period of time, the disease is most commonly manifested as an acute inflammation of the lungs, resulting in fever and a nonproductive cough. If left untreated, serious complications and death typically result.

Although tuberculosis can generally be controlled using extended antibiotic therapy, such treatment is not sufficient to prevent the spread of the disease. Infected individuals may be asymptomatic, but contagious, for some time. In addition, although compliance with the treatment regimen is critical, patient behavior is difficult to monitor. Some patients do not complete the course of treatment, which can lead to ineffective treatment and the development of drug resistance.

Inhibiting the spread of tuberculosis will require effective vaccination and accurate, early diagnosis of the disease. Currently, vaccination with live bacteria is the most efficient method for inducing protective immunity. The most common Mycobacterium for this purpose is Bacillus Calmette-Guerin (BCG), an avirulent strain of *Mycobacterium bovis*.

However, the safety and efficacy of BCG is a source of controversy and some countries, such as the United States, do not vaccinate the general public. Diagnosis is commonly achieved using a skin test, which involves intradermal exposure to tuberculin PPD (protein-purified derivative). Antigen-specific T cell responses result in measurable incubation at the injection site by 48-72 hours after injection, which indicates exposure to Mycobacterial antigens. Sensitivity and specificity have, however, been a problem with this test, and individuals vaccinated with BCG cannot be distinguished from infected individuals.

While macrophages have been shown to act as the principal effectors of *M. tuberculosis* immunity, T cells are the predominant inducers of such immunity. The essential role of T cells in protection against *M. tuberculosis* infection is illustrated by the frequent occurrence of *M. tuberculosis* in AIDS patients, due to the depletion of CD4 T cells associated with human immunodeficiency virus (HIV) infection. Mycobacterium-reactive CD4 T cells have been shown to be potent producers of gamma-interferon (IFN-γ), which, in turn, has been shown to trigger the anti-mycobacterial effects of macrophages in mice. While the role of IFN-γ in humans is less clear, studies have shown that 1,25-dihydroxy-vitamin D3, either alone or in combination with IFN-γ or tumor necrosis factor-alpha, activates human macrophages to inhibit *M. tuberculosis* infection. Furthermore, it is known that IFN-γ stimulates human macrophages to make 1,25-dihydroxy-vitamin D3. Similarly, IL-12 has been shown to play a role in stimulating resistance to *M. tuberculosis* infection. For a review of the immunology of *M. tuberculosis* infection *see* Chan and Kaufmann, in *Tuberculosis: Pathogenesis, Protection and Control*, Bloom (ed.), ASM Press, Washington, DC, 1994.

Accordingly, there is a need in the art for improved diagnostic methods for detecting tuberculosis. The present invention fulfills this need and further provides other related advantages.

Summary of the Invention

Briefly stated, the present invention provides compositions and methods for diagnosing tuberculosis.

In one embodiment, polypeptides are provided that comprise an antigenic portion of a *M. tuberculosis* antigen, or a variant of such an antigen that differs only in

conservative substitutions and/or modifications, wherein the antigen comprises an amino acid sequence encoded by a DNA sequence selected from the group consisting of the sequences recited in SEQ ID NO: 1, 11, 12, 83, 103-108, 125, 127, 129-137, 139 and 140, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID NO: 1, 11, 12, 83, 103-108, 125, 127, 129-137, 139 and 140, or a complement thereof, under moderately stringent conditions. In a second embodiment, the present invention provides polypeptides comprising an immunogenic portion of a *M. tuberculosis* antigen having an amino acid sequence selected from the group consisting of sequences provided in SEQ ID NO: 16-33, 109, 126, 138, 141, 142 and variants thereof

In related aspects, DNA sequences encoding the above polypeptides, recombinant expression vectors comprising these DNA sequences and host cells transformed or transfected with such expression vectors are also provided.

In another aspect, the present invention provides fusion proteins comprising a first and a second inventive polypeptide or, alternatively, an inventive polypeptide and a known *M. tuberculosis* antigen.

In further aspects of the subject invention, methods and diagnostic kits are provided for detecting tuberculosis in a patient. The methods comprise: (a) contacting a biological sample with at least one of the above polypeptides; and (b) detecting in the sample the presence of antibodies that bind to the polypeptide or polypeptides, thereby detecting *M. tuberculosis* infection in the biological sample. Suitable biological samples include whole blood, sputum, serum, plasma, saliva, cerebrospinal fluid and urine. The diagnostic kits comprise one or more of the above polypeptides in combination with a detection reagent.

The present invention also provides methods for detecting *M. tuberculosis* infection comprising: (a) obtaining a biological sample from a patient; (b) contacting the sample with at least one oligonucleotide primer in a polymerase chain reaction, the oligonucleotide primer being specific for a DNA sequence encoding the above polypeptides; and (c) detecting in the sample a DNA sequence that amplifies in the presence of the first and second oligonucleotide primers. In one embodiment, the oligonucleotide primer comprises at least about 10 contiguous nucleotides of such a DNA sequence.

In a further aspect, the present invention provides a method for detecting *M. tuberculosis* infection in a patient comprising: (a) obtaining a biological sample from the patient; (b) contacting the sample with an oligonucleotide probe specific for a DNA sequence encoding the above polypeptides; and (c) detecting in the sample a DNA sequence that hybridizes to the oligonucleotide probe. In one embodiment, the oligonucleotide probe comprises at least about 15 contiguous nucleotides of such a DNA sequence.

In yet another aspect, methods are provided for detecting tuberculosis in a patient, such methods comprising contacting a biological sample with one or more polypeptides encoded by a DNA sequence selected from the group consisting of SEQ ID NO: 2-10, 102, 128, the complements of said sequences, and DNA sequences that hybridize to a sequence recited in SEQ ID NO: 2-10, 102, 128; and detecting in the sample the presence of antibodies that bind to the polypeptide, thereby detecting *M. tuberculosis* infection in the biological sample. Diagnostic kits for use in such methods are also provided.

In another aspect, the present invention provides antibodies, both polyclonal and monoclonal, that bind to the polypeptides described above, as well as methods for their use in the detection of *M. tuberculosis* infection.

These and other aspects of the present invention will become apparent upon reference to the following detailed description and attached drawings. All references disclosed herein are hereby incorporated by reference in their entirety as if each was incorporated individually.

Brief Description of the Drawings

Figures 1 and 1B illustrate the stimulation of proliferation and interferon-γ production, respectively, in T cells derived from a first PPD-positive donor (referred to as D7) by recombinant ORF-2 and synthetic peptides to ORF-2.

Figures 2A and 2B illustrate the stimulation of proliferation and interferon-γ production, respectively, in T cells derived from a second PPD-positive donor (referred to as D160) by recombinant ORF-2 and synthetic peptides to ORF-2.

Detailed Description of the Invention

As noted above, the present invention is generally directed to compositions and methods for preventing, treating and diagnosing tuberculosis. The compositions of the subject invention include polypeptides that comprise at least one immunogenic portion of a *M. tuberculosis* antigen, or a variant of such an antigen that differs only in conservative substitutions and/or modifications. As used herein, the term "polypeptide" encompasses amino acid chains of any length, including full length proteins (*i.e.*, antigens), wherein the amino acid residues are linked by covalent peptide bonds. Thus, a polypeptide comprising an immunogenic portion of one of the above antigens may consist entirely of the immunogenic portion, or may contain additional sequences. The additional sequences may be derived from the native *M. tuberculosis* antigen or may be heterologous, and such sequences may (but need not) be immunogenic.

"Immunogenic," as used herein, refers to the ability to elicit an immune response (*e.g.*, cellular) in a patient, such as a human, and/or in a biological sample. In particular, antigens that are immunogenic (and immunogenic portions or other variants of such antigens) are capable of stimulating cell proliferation, interleukin-12 production and/or interferon-γ production in biological samples comprising one or more cells selected from the group of T cells, NK cells, B cells and macrophages, where the cells are derived from an *M. tuberculosis*-immune individual. Polypeptides comprising at least an immunogenic portion of one or more *M. tuberculosis* antigens may generally be used to detect tuberculosis or to induce protective immunity against tuberculosis in a patient.

The compositions and methods of this invention also encompass variants of the above polypeptides. A polypeptide "variant," as used herein, is a polypeptide that differs from the recited polypeptide only in conservative substitutions and/or modifications, such that the therapeutic, antigenic and/or immunogenic properties of the polypeptide are retained. Polypeptide variants preferably exhibit at least about 70%, more preferably at least about 90% and most preferably at least about 95% identity to the identified polypeptides. For polypeptides with immunoreactive properties, variants may, alternatively, be identified by modifying the amino acid sequence of one of the above polypeptides, and evaluating the immunoreactivity of the modified polypeptide. For polypeptides useful for the generation of

diagnostic binding agents, a variant may be identified by evaluating a modified polypeptide for the ability to generate antibodies that detect the presence or absence of tuberculosis. Alternatively, variants of the claimed antigens that may be usefully employed in the inventive diagnostic methods may be identified by evaluating modified polypeptides for their ability to detect antibodies present in the sera of tuberculosis-infected patients. Such modified sequences may be prepared and tested using, for example, the representative procedures described herein.

A "conservative substitution" is one in which an amino acid is substituted for another amino acid that has similar properties, such that one skilled in the art of peptide chemistry would expect the secondary structure and hydropathic nature of the polypeptide to be substantially unchanged. In general, the following groups of amino acids represent conservative changes: (1) ala, pro, gly, glu, asp, gln, asn, ser, thr; (2) cys, ser, tyr, thr; (3) val, ile, leu, met, ala, phe; (4) lys, arg, his; and (5) phe, tyr, trp, his.

Variants may also (or alternatively) be modified by, for example, the deletion or addition of amino acids that have minimal influence on the immunogenic properties, secondary structure and hydropathic nature of the polypeptide. For example, a polypeptide may be conjugated to a signal (or leader) sequence at the N-terminal end of the protein which co-translationally or post-translationally directs transfer of the protein. The polypeptide may also be conjugated to a linker or other sequence for ease of synthesis, purification or identification of the polypeptide (*e.g.*, poly-His), or to enhance binding of the polypeptide to a solid support. For example, a polypeptide may be conjugated to an immunoglobulin Fc region.

In general, *M. tuberculosis* antigens, and DNA sequences encoding such antigens, may be prepared using any of a variety of procedures. For example, genomic or cDNA libraries derived from *M. tuberculosis* may be screened directly using peripheral blood mononuclear cells (PBMCs) or T cell lines or clones derived from one or more *M. tuberculosis*-immune individuals. Direct library screens may generally be performed by assaying pools of expressed recombinant proteins for the ability of induce proliferation and/or interferon-γ production in T cells derived from an *M. tuberculosis*-immune individual.

Potential T cell antigens may be first selected based on antibody reactivity, as described above.

Alternatively, DNA sequences encoding antigens may be identified by screening an appropriate *M. tuberculosis* genomic or cDNA expression library with sera obtained from patients infected with *M. tuberculosis*. Such screens may generally be performed using techniques well known to those of ordinary skill in the art, such as those described in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989.

Purified antigens are then evaluated for their ability to elicit an appropriate immune response (*e.g.*, cellular) using, for example, the representative methods described herein. Immunogenic antigens may then be partially sequenced using techniques such as traditional Edman chemistry. *See* Edman and Berg, *Eur. J. Biochem. 80*:116-132, 1967. Immunogenic antigens may also be produced recombinantly using a DNA sequence that encodes the antigen, which has been inserted into an expression vector and expressed in an appropriate host.

DNA sequences encoding the inventive antigens may also be obtained by screening an appropriate *M. tuberculosis* cDNA or genomic DNA library for DNA sequences that hybridize to degenerate oligonucleotides derived from partial amino acid sequences of isolated antigens. Degenerate oligonucleotide sequences for use in such a screen may be designed and synthesized, and the screen may be performed, as described (for example) in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989 (and references cited therein). Polymerase chain reaction (PCR) may also be employed, using the above oligonucleotides in methods well known in the art, to isolate a nucleic acid probe from a cDNA or genomic library. The library screen may then be performed using the isolated probe.

Regardless of the method of preparation, the antigens described herein are "antigenic." More specifically, the antigens have the ability to react with sera obtained from an *M. tuberculosis*-infected individual. Reactivity may be evaluated using, for example, the representative ELISA assays described herein, where an absorbance reading with sera from

infected individuals that is at least three standard deviations above the absorbance obtained with sera from uninfected individuals is considered positive.

Antigenic portions of *M. tuberculosis* antigens may be prepared and identified using well known techniques, such as those summarized in Paul, *Fundamental Immunology*, 3d ed., Raven Press, 1993, pp. 243-247 and references cited therein. Such techniques include screening polypeptide portions of the native antigen for antigenic properties. The representative ELISAs described herein may generally be employed in these screens. An antigenic portion of a polypeptide is a portion that, within such representative assays, generates a signal in such assays that is substantially similar to that generated by the full length antigen. In other words, an antigenic portion of a *M. tuberculosis* antigen generates at least about 20%, and preferably about 100%, of the signal induced by the full length antigen in a model ELISA as described herein.

Portions and other variants of *M. tuberculosis* antigens may be generated by synthetic or recombinant means. Synthetic polypeptides having fewer than about 100 amino acids, and generally fewer than about 50 amino acids, may be generated using techniques well known in the art. For example, such polypeptides may be synthesized using any of the commercially available solid-phase techniques, such as the Merrifield solid-phase synthesis method, where amino acids are sequentially added to a growing amino acid chain. *See* Merrifield, *J. Am. Chem. Soc.* 85:2149-2146, 1963. Equipment for automated synthesis of polypeptides is commercially available from suppliers such as Applied BioSystems, Inc., Foster City, CA, and may be operated according to the manufacturer's instructions. Variants of a native antigen may generally be prepared using standard mutagenesis techniques, such as oligonucleotide-directed site-specific mutagenesis. Sections of the DNA sequence may also be removed using standard techniques to permit preparation of truncated polypeptides.

Recombinant polypeptides containing portions and/or variants of a native antigen may be readily prepared from a DNA sequence encoding the polypeptide using a variety of techniques well known to those of ordinary skill in the art. For example, supernatants from suitable host/vector systems which secrete recombinant protein into culture media may be first concentrated using a commercially available filter. Following concentration, the concentrate may be applied to a suitable purification matrix such as an

affinity matrix or an ion exchange resin. Finally, one or more reverse phase HPLC steps can be employed to further purify a recombinant protein.

Any of a variety of expression vectors known to those of ordinary skill in the art may be employed to express recombinant polypeptides as described herein. Expression may be achieved in any appropriate host cell that has been transformed or transfected with an expression vector containing a DNA molecule that encodes a recombinant polypeptide. Suitable host cells include prokaryotes, yeast and higher eukaryotic cells. Preferably, the host cells employed are *E. coli*, yeast or a mammalian cell line, such as COS or CHO. The DNA sequences expressed in this manner may encode naturally occurring antigens, portions of naturally occurring antigens, or other variants thereof.

In general, regardless of the method of preparation, the polypeptides disclosed herein are prepared in substantially pure form. Preferably, the polypeptides are at least about 80% pure, more preferably at least about 90% pure and most preferably at least about 99% pure. For use in the methods described herein, however, such substantially pure polypeptides may be combined.

In one embodiment, the subject invention discloses polypeptides comprising at least an antigenic portion of a *M. tuberculosis* antigen (or a variant of such an antigen) that comprises one or more of the amino acid sequences encoded by (a) the DNA sequences of SEQ ID NO: 1-12, 83, 102-108, 125, 127-137, 139 and 140, (b) the complements of such DNA sequences or (c) DNA sequences substantially homologous to a sequence in (a) or (b). In a related embodiment, the present invention provides polypeptides comprising at least an immunogenic portion of an *M. tuberculosis* antigen having an amino acid sequence selected from the group consisting of sequences provided in SEQ ID NO: 16-33, 109, 126, 138, 141, 142 and variants thereof.

The *M. tuberculosis* antigens provided herein include variants that are encoded by DNA sequences which are substantially homologous to one or more of DNA sequences specifically recited herein. "Substantial homology," as used herein, refers to DNA sequences that are capable of hybridizing under moderately stringent conditions. Suitable moderately stringent conditions include prewashing in a solution of 5X SSC, 0.5% SDS, 1.0 mM EDTA (pH 8.0); hybridizing at 50°C-65°C, 5X SSC, overnight or, in the event of cross-species

homology, at 45°C with 0.5X SSC; followed by washing twice at 65°C for 20 minutes with each of 2X, 0.5X and 0.2X SSC containing 0.1% SDS). Such hybridizing DNA sequences are also within the scope of this invention, as are nucleotide sequences that, due to code degeneracy, encode an immunogenic polypeptide that is encoded by a hybridizing DNA sequence.

In a related aspect, the present invention provides fusion proteins comprising a first and a second inventive polypeptide or, alternatively, a polypeptide of the present invention and a known *M. tuberculosis* antigen, such as the 38 kD antigen described in Andersen and Hansen, *Infect. Immun. 57*:2481-2488, 1989, (Genbank Accession No. M30046), or ESAT-6 previously identified in *M. bovis* (Accession No. U34848) and in *M. tuberculosis* (Sorensen et al., *Infec. Immun. 63*:1710-1717, 1995). Variants of such fusion proteins are also provided. The fusion proteins of the present invention may include a linker peptide between the first and second polypeptides.

A DNA sequence encoding a fusion protein of the present invention is constructed using known recombinant DNA techniques to assemble separate DNA sequences encoding the first and second polypeptides into an appropriate expression vector. The 3' end of a DNA sequence encoding the first polypeptide is ligated, with or without a peptide linker, to the 5' end of a DNA sequence encoding the second polypeptide so that the reading frames of the sequences are in phase to permit mRNA translation of the two DNA sequences into a single fusion protein that retains the biological activity of both the first and the second polypeptides.

A peptide linker sequence may be employed to separate the first and the second polypeptides by a distance sufficient to ensure that each polypeptide folds into its secondary and tertiary structures. Such a peptide linker sequence is incorporated into the fusion protein using standard techniques well known in the art. Suitable peptide linker sequences may be chosen based on the following factors: (1) their ability to adopt a flexible extended conformation; (2) their inability to adopt a secondary structure that could interact with functional epitopes on the first and second polypeptides; and (3) the lack of hydrophobic or charged residues that might react with the polypeptide functional epitopes. Preferred peptide linker sequences contain Gly, Asn and Ser residues. Other near neutral amino acids,

such as Thr and Ala may also be used in the linker sequence. Amino acid sequences which may be usefully employed as linkers include those disclosed in Maratea et al., *Gene 40*:39-46, 1985; Murphy et al., *Proc. Natl. Acad. Sci. USA 83*:8258-8562, 1986; U.S. Patent No. 4,935,233 and U.S. Patent No. 4,751,180. The linker sequence may be from 1 to about 50 amino acids in length. Peptide linker sequences are not required when the first and second polypeptides have non-essential N-terminal amino acid regions that can be used to separate the functional domains and prevent steric hindrance.

In another aspect, the present invention provides methods for using the polypeptides described above to diagnose tuberculosis. In this aspect, methods are provided for detecting *M. tuberculosis* infection in a biological sample, using one or more of the above polypeptides, alone or in combination. In embodiments in which multiple polypeptides are employed, polypeptides other than those specifically described herein, such as the 38 kD antigen described above, may be included. As used herein, a "biological sample" is any antibody-containing sample obtained from a patient. Preferably, the sample is whole blood, sputum, serum, plasma, saliva, cerebrospinal fluid or urine. More preferably, the sample is a blood, serum or plasma sample obtained from a patient or a blood supply. The polypeptide(s) are used in an assay, as described below, to determine the presence or absence of antibodies to the polypeptide(s) in the sample, relative to a predetermined cut-off value. The presence of such antibodies indicates previous sensitization to mycobacterial antigens which may be indicative of tuberculosis.

In embodiments in which more than one polypeptide is employed, the polypeptides used are preferably complementary (*i.e.*, one component polypeptide will tend to detect infection in samples where the infection would not be detected by another component polypeptide). Complementary polypeptides may generally be identified by using each polypeptide individually to evaluate serum samples obtained from a series of patients known to be infected with *M. tuberculosis*. After determining which samples test positive (as described below) with each polypeptide, combinations of two or more polypeptides may be formulated that are capable of detecting infection in most, or all, of the samples tested. Such polypeptides are complementary. For example, approximately 25-30% of sera from tuberculosis-infected individuals are negative for antibodies to any single protein, such as the

38 kD antigen mentioned above. Complementary polypeptides may, therefore, be used in combination with the 38 kD antigen to improve sensitivity of a diagnostic test.

There are a variety of assay formats known to those of ordinary skill in the art for using one or more polypeptides to detect antibodies in a sample. *See, e.g.*, Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988, which is incorporated herein by reference. In a preferred embodiment, the assay involves the use of polypeptide immobilized on a solid support to bind to and remove the antibody from the sample. The bound antibody may then be detected using a detection reagent that contains a reporter group. Suitable detection reagents include antibodies that bind to the antibody/polypeptide complex and free polypeptide labeled with a reporter group (*e.g.*, in a semi-competitive assay). Alternatively, a competitive assay may be utilized, in which an antibody that binds to the polypeptide is labeled with a reporter group and allowed to bind to the immobilized antigen after incubation of the antigen with the sample. The extent to which components of the sample inhibit the binding of the labeled antibody to the polypeptide is indicative of the reactivity of the sample with the immobilized polypeptide.

The solid support may be any solid material known to those of ordinary skill in the art to which the antigen may be attached. For example, the solid support may be a test well in a microtiter plate or a nitrocellulose or other suitable membrane. Alternatively, the support may be a bead or disc, such as glass, fiberglass, latex or a plastic material such as polystyrene or polyvinylchloride. The support may also be a magnetic particle or a fiber optic sensor, such as those disclosed, for example, in U.S. Patent No. 5,359,681.

The polypeptides may be bound to the solid support using a variety of techniques known to those of ordinary skill in the art, which are amply described in the patent and scientific literature. In the context of the present invention, the term "bound" refers to both noncovalent association, such as adsorption, and covalent attachment (which may be a direct linkage between the antigen and functional groups on the support or may be a linkage by way of a cross-linking agent). Binding by adsorption to a well in a microtiter plate or to a membrane is preferred. In such cases, adsorption may be achieved by contacting the polypeptide, in a suitable buffer, with the solid support for a suitable amount of time. The contact time varies with temperature, but is typically between about 1 hour and 1 day. In

general, contacting a well of a plastic microtiter plate (such as polystyrene or polyvinylchloride) with an amount of polypeptide ranging from about 10 ng to about 1 μg, and preferably about 100 ng, is sufficient to bind an adequate amount of antigen.

Covalent attachment of polypeptide to a solid support may generally be achieved by first reacting the support with a bifunctional reagent that will react with both the support and a functional group, such as a hydroxyl or amino group, on the polypeptide. For example, the polypeptide may be bound to supports having an appropriate polymer coating using benzoquinone or by condensation of an aldehyde group on the support with an amine and an active hydrogen on the polypeptide (*see*, *e.g.*, Pierce Immunotechnology Catalog and Handbook, 1991, at A12-A13).

In certain embodiments, the assay is an enzyme linked immunosorbent assay (ELISA). This assay may be performed by first contacting a polypeptide antigen that has been immobilized on a solid support, commonly the well of a microtiter plate, with the sample, such that antibodies to the polypeptide within the sample are allowed to bind to the immobilized polypeptide. Unbound sample is then removed from the immobilized polypeptide and a detection reagent capable of binding to the immobilized antibody-polypeptide complex is added. The amount of detection reagent that remains bound to the solid support is then determined using a method appropriate for the specific detection reagent.

More specifically, once the polypeptide is immobilized on the support as described above, the remaining protein binding sites on the support are typically blocked. Any suitable blocking agent known to those of ordinary skill in the art, such as bovine serum albumin or Tween 20™ (Sigma Chemical Co., St. Louis, MO) may be employed. The immobilized polypeptide is then incubated with the sample, and antibody is allowed to bind to the antigen. The sample may be diluted with a suitable diluent, such as phosphate-buffered saline (PBS) prior to incubation. In general, an appropriate contact time (*i.e.*, incubation time) is that period of time that is sufficient to detect the presence of antibody within a *M. tuberculosis*-infected sample. Preferably, the contact time is sufficient to achieve a level of binding that is at least 95% of that achieved at equilibrium between bound and unbound antibody. Those of ordinary skill in the art will recognize that the time necessary to achieve equilibrium may be readily determined by assaying the level of binding that occurs over a

period of time. At room temperature, an incubation time of about 30 minutes is generally sufficient.

Unbound sample may then be removed by washing the solid support with an appropriate buffer, such as PBS containing 0.1% Tween 20™. Detection reagent may then be added to the solid support. An appropriate detection reagent is any compound that binds to the immobilized antibody-polypeptide complex and that can be detected by any of a variety of means known to those in the art. Preferably, the detection reagent contains a binding agent (such as, for example, Protein A, Protein G, immunoglobulin, lectin or free antigen) conjugated to a reporter group. Preferred reporter groups include enzymes (such as horseradish peroxidase), substrates, cofactors, inhibitors, dyes, radionuclides, luminescent groups, fluorescent groups and biotin. The conjugation of binding agent to reporter group may be achieved using standard methods known to those of ordinary skill in the art. Common binding agents may also be purchased conjugated to a variety of reporter groups from many commercial sources (*e.g.*, Zymed Laboratories, San Francisco, CA, and Pierce, Rockford, IL).

The detection reagent is then incubated with the immobilized antibody-polypeptide complex for an amount of time sufficient to detect the bound antibody. An appropriate amount of time may generally be determined from the manufacturer's instructions or by assaying the level of binding that occurs over a period of time. Unbound detection reagent is then removed and bound detection reagent is detected using the reporter group. The method employed for detecting the reporter group depends upon the nature of the reporter group. For radioactive groups, scintillation counting or autoradiographic methods are generally appropriate. Spectroscopic methods may be used to detect dyes, luminescent groups and fluorescent groups. Biotin may be detected using avidin, coupled to a different reporter group (commonly a radioactive or fluorescent group or an enzyme). Enzyme reporter groups may generally be detected by the addition of substrate (generally for a specific period of time), followed by spectroscopic or other analysis of the reaction products.

To determine the presence or absence of anti-*M. tuberculosis* antibodies in the sample, the signal detected from the reporter group that remains bound to the solid support is generally compared to a signal that corresponds to a predetermined cut-off value. In one

preferred embodiment, the cut-off value is the average mean signal obtained when the immobilized antigen is incubated with samples from an uninfected patient. In general, a sample generating a signal that is three standard deviations above the predetermined cut-off value is considered positive for tuberculosis. In an alternate preferred embodiment, the cut-off value is determined using a Receiver Operator Curve, according to the method of Sackett et al., *Clinical Epidemiology: A Basic Science for Clinical Medicine*, Little Brown and Co., 1985, pp. 106-107. Briefly, in this embodiment, the cut-off value may be determined from a plot of pairs of true positive rates (*i.e.*, sensitivity) and false positive rates (100%-specificity) that correspond to each possible cut-off value for the diagnostic test result. The cut-off value on the plot that is the closest to the upper left-hand corner (*i.e.*, the value that encloses the largest area) is the most accurate cut-off value, and a sample generating a signal that is higher than the cut-off value determined by this method may be considered positive. Alternatively, the cut-off value may be shifted to the left along the plot, to minimize the false positive rate, or to the right, to minimize the false negative rate. In general, a sample generating a signal that is higher than the cut-off value determined by this method is considered positive for tuberculosis.

In a related embodiment, the assay is performed in a rapid flow-through or strip test format, wherein the antigen is immobilized on a membrane, such as nitrocellulose. In the flow-through test, antibodies within the sample bind to the immobilized polypeptide as the sample passes through the membrane. A detection reagent (*e.g.*, protein A-colloidal gold) then binds to the antibody-polypeptide complex as the solution containing the detection reagent flows through the membrane. The detection of bound detection reagent may then be performed as described above. In the strip test format, one end of the membrane to which polypeptide is bound is immersed in a solution containing the sample. The sample migrates along the membrane through a region containing detection reagent and to the area of immobilized polypeptide. Concentration of detection reagent at the polypeptide indicates the presence of anti-*M. tuberculosis* antibodies in the sample. Typically, the concentration of detection reagent at that site generates a pattern, such as a line, that can be read visually. The absence of such a pattern indicates a negative result. In general, the amount of polypeptide immobilized on the membrane is selected to generate a visually discernible pattern when the

biological sample contains a level of antibodies that would be sufficient to generate a positive signal in an ELISA, as discussed above. Preferably, the amount of polypeptide immobilized on the membrane ranges from about 25 ng to about 1 μg, and more preferably from about 50 ng to about 500 ng. Such tests can typically be performed with a very small amount (*e.g.*, one drop) of patient serum or blood.

Of course, numerous other assay protocols exist that are suitable for use with the polypeptides of the present invention. The above descriptions are intended to be exemplary only.

In yet another aspect, the present invention provides antibodies to the inventive polypeptides. Antibodies may be prepared by any of a variety of techniques known to those of ordinary skill in the art. *See, e.g.*, Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988. In one such technique, an immunogen comprising the antigenic polypeptide is initially injected into any of a wide variety of mammals (*e.g.*, mice, rats, rabbits, sheep and goats). In this step, the polypeptides of this invention may serve as the immunogen without modification. Alternatively, particularly for relatively short polypeptides, a superior immune response may be elicited if the polypeptide is joined to a carrier protein, such as bovine serum albumin or keyhole limpet hemocyanin. The immunogen is injected into the animal host, preferably according to a predetermined schedule incorporating one or more booster immunizations, and the animals are bled periodically. Polyclonal antibodies specific for the polypeptide may then be purified from such antisera by, for example, affinity chromatography using the polypeptide coupled to a suitable solid support.

Monoclonal antibodies specific for the antigenic polypeptide of interest may be prepared, for example, using the technique of Kohler and Milstein, *Eur. J. Immunol.* 6:511-519, 1976, and improvements thereto. Briefly, these methods involve the preparation of immortal cell lines capable of producing antibodies having the desired specificity (*i.e.*, reactivity with the polypeptide of interest). Such cell lines may be produced, for example, from spleen cells obtained from an animal immunized as described above. The spleen cells are then immortalized by, for example, fusion with a myeloma cell fusion partner, preferably one that is syngeneic with the immunized animal. A variety of fusion techniques may be

employed. For example, the spleen cells and myeloma cells may be combined with a nonionic detergent for a few minutes and then plated at low density on a selective medium that supports the growth of hybrid cells, but not myeloma cells. A preferred selection technique uses HAT (hypoxanthine, aminopterin, thymidine) selection. After a sufficient time, usually about 1 to 2 weeks, colonies of hybrids are observed. Single colonies are selected and tested for binding activity against the polypeptide. Hybridomas having high reactivity and specificity are preferred.

Monoclonal antibodies may be isolated from the supernatants of growing hybridoma colonies. In addition, various techniques may be employed to enhance the yield, such as injection of the hybridoma cell line into the peritoneal cavity of a suitable vertebrate host, such as a mouse. Monoclonal antibodies may then be harvested from the ascites fluid or the blood. Contaminants may be removed from the antibodies by conventional techniques, such as chromatography, gel filtration, precipitation, and extraction. The polypeptides of this invention may be used in the purification process in, for example, an affinity chromatography step.

Antibodies may be used in diagnostic tests to detect the presence of *M. tuberculosis* antigens using assays similar to those detailed above and other techniques well known to those of skill in the art, thereby providing a method for detecting *M. tuberculosis* infection in a patient.

Diagnostic reagents of the present invention may also comprise DNA sequences encoding one or more of the above polypeptides, or one or more portions thereof. For example, at least two oligonucleotide primers may be employed in a polymerase chain reaction (PCR) based assay to amplify *M. tuberculosis*-specific cDNA derived from a biological sample, wherein at least one of the oligonucleotide primers is specific for a DNA molecule encoding a polypeptide of the present invention. The presence of the amplified cDNA is then detected using techniques well known in the art, such as gel electrophoresis. Similarly, oligonucleotide probes specific for a DNA molecule encoding a polypeptide of the present invention may be used in a hybridization assay to detect the presence of an inventive polypeptide in a biological sample.

As used herein, the term "oligonucleotide primer/probe specific for a DNA molecule" means an oligonucleotide sequence that has at least about 80%, preferably at least about 90% and more preferably at least about 95%, identity to the DNA molecule in question. Oligonucleotide primers and/or probes which may be usefully employed in the inventive diagnostic methods preferably have at least about 10-40 nucleotides. In a preferred embodiment, the oligonucleotide primers comprise at least about 10 contiguous nucleotides of a DNA molecule encoding one of the polypeptides disclosed herein. Preferably, oligonucleotide probes for use in the inventive diagnostic methods comprise at least about 15 contiguous oligonucleotides of a DNA molecule encoding one of the polypeptides disclosed herein. Techniques for both PCR based assays and hybridization assays are well known in the art (see, for example, Mullis *et al. Ibid*; Ehrlich, *Ibid*). Primers or probes may thus be used to detect *M. tuberculosis*-specific sequences in biological samples. DNA probes or primers comprising oligonucleotide sequences described above may be used alone, in combination with each other, or with previously identified sequences, such as the 38 kD antigen discussed above.

The following Examples are offered by way of illustration and not by way of limitation.

EXAMPLES

EXAMPLE 1

PURIFICATION AND CHARACTERIZATION OF *M. TUBERCULOSIS* POLYPEPTIDES USING CD4+ T CELL LINES GENERATED FROM HUMAN PBMC

M. tuberculosis antigens of the present invention were isolated by expression cloning of cDNA libraries of *M. tuberculosis* strains H37Rv and Erdman essentially as described by Sanderson et al. (*J. Exp. Med.*, 1995, *182*:1751-1757) and were shown to induce PBMC proliferation and IFN-γ in an immunoreactive T cell line.

Two CD4+ T cell lines, referred to as DC-4 and DC-5, were generated against dendritic cells infected with *M. tuberculosis*. Specifically, dendritic cells were prepared from adherent PBMC from a single donor and subsequently infected with tuberculosis. Lymphocytes from the same donor were cultured under limiting dilution conditions with the infected dendritic cells to generate the CD4+ T cell lines DC-4 and DC-5. These cell lines were shown to react with crude soluble proteins from *M. tuberculosis* but not with Tb38-1. Limiting dilution conditions were employed to obtain a third CD4+ T cell line, referred to as DC-6, which was shown to react with both crude soluble proteins and Tb38-1.

Genomic DNA was isolated from the *M. tuberculosis* strains H37Rv and Erdman and used to construct expression libraries in the vector pBSK(-)using the Lambda ZAP expression system (Stratagene, La Jolla, CA). These libraries were transformed into *E. coli*, pools of induced *E. coli* cultures were incubated with dendritic cells, and the ability of the resulting incubated dendritic cells to stimulate cell proliferation and IFN-γ production in the CD4+ T cell line DC-6 was examined as described below in Example 2. Positive pools were fractionated and re-tested until pure *M. tuberculosis* clones were obtained.

Nineteen clones were isolated, of which nine were found to contain the previously identified *M. tuberculosis* antigens TbH-9 and Tb38-1, disclosed in U.S. Patent Application Nos. 08/533,634. The determined cDNA sequences for the remaining ten clones (hereinafter referred to as Tb224, Tb636, Tb424, Tb436, Tb398, Tb508, Tb441, Tb475, Tb488 and Tb465) are provided in SEQ ID No: 1-10, respectively. The corresponding predicted amino acid sequences for Tb224 and Tb636 are provided in SEQ ID NO: 13 and 14, respectively. The open reading frames for these two antigens were found to show some homology to TbH-9. Tb224 and Tb636 were also found to be overlapping clones.

Tb424, Tb436, Tb398, Tb508, Tb441, Tb475, Tb488 and Tb465 were each found to contain two small open reading frames (referred to as ORF-1 and ORF-2) or truncated forms thereof, with minor variations in ORF-1 and ORF-2 being found for each clone. The predicted amino acid sequences of ORF-1 and ORF-2 for Tb424, Tb436, Tb398, Tb508, Tb441, Tb475, Tb488 and Tb465 are provided in SEQ ID NO: 16 and 17, 18 and 19, 20 and 21, 22 and 23, 24 and 25, 26 and 27, 28 and 29, and 30 and 31, respectively. In addition, clones Tb424 and Tb436 were found to contain a third apparent open reading frame, referred to as ORF-U. The predicted amino acid sequences of ORF-U for Tb424 and Tb436 are provided in SEQ ID NO: 32 and 33, respectively. Tb424 and Tb436 were found to be either overlapping clones or recently duplicated/transposed copies. Similarly Tb398, Tb508 and Tb465 were found to be either overlapping clones or recently duplicated/transposed copies, as were Tb475 and Tb488.

These sequences were compared with known sequences in the gene bank using the BLASTN system. No homologies to the antigens Tb224 and Tb431 were found. Tb636 was found to be 100% identical to a cosmid previously identified in *M. tuberculosis*. Similarly, Tb508, Tb488, Tb398, Tb424, Tb436, Tb441, Tb465 and Tb475 were found to show homology to known *M. tuberculosis* cosmids. In addition, Tb488 was found to have 100% homology to *M. tuberculosis* topoisomerase I.

Seventeen overlapping peptides to the open reading frames ORF-1 (referred to as 1-1 - 1-17; SEQ ID NO: 34-50, respectively) and thirty overlapping peptides to the open reading frame ORF-2 (referred to as 2-1 - 2-30, SEQ ID NO: 51-80) were synthesized using the procedure described below in Example 3.

The ability of the synthetic peptides, and of recombinant ORF-1 and ORF-2, to induce T cell proliferation and IFN-γ production in PBMC from PPD-positive donors was assayed as described below in Example 2. Figs. 1A-B and 2A-B illustrate stimulation of T cell proliferation and IFN-γ by recombinant ORF-2 and the synthetic peptides 2-1 - 2-16 for two donors, referred to as D7 and D160, respectively. Recombinant ORF-2 (referred to as MTI) stimulated T cell proliferation and IFN-γ production in PBMC from both donors. The amount of PBMC stimulation seen with the individual synthetic peptides varied with each donor, indicating that each donor recognizes different epitopes on ORF2. The proteins encoded by ORF-1, ORF-2 and ORF-U were subsequently named MTS, MTI and MSF, respectively.

Eighteen overlapping peptides to the sequence of MSF (referred to as MSF-1 – MSF-18; SEQ ID NO: 84-101, respectively) were synthesized and their ability to stimulate T cell proliferation and IFN-γ production in a CD4+ T cell line generated against *M. tuberculosis* culture filtrate was examined as described below. The peptides referred to as MSF-12 and MSF-13 (SEQ ID NO: 95 and 96, respectively) were found to show the highest levels of reactivity.

Two overlapping peptides (SEQ ID NO:81 and 82) to the open reading frame of Tb224 were synthesized and shown to induce T cell proliferation and IFN-γ production in PBMC from PPD-positive donors.

Two CD4+ T cell lines from different donors were generated against *M. tuberculosis* infected dendritic cells using the above methodology. Screening of the *M. tuberculosis* cDNA expression library described above using this cell line, resulted in the isolation of two clones referred to as Tb867 and Tb391. The determined cDNA sequence for Tb867 (SEQ ID NO: 102) was found to be identical to the previously isolated *M. tuberculosis* cosmid SCY22G10, with the candidate reactive open reading frame encoding a 750 amino acid *M. tuberculosis* protein kinase. Comparison of the determined cDNA sequence for Tb391 (SEQ ID NO: 103) with those in the gene bank revealed no significant homologies to known sequences.

In further studies, CD4+ T cell lines were generated against *M. tuberculosis* culture filtrate, essentially as outlined above, and used to screen the *M. tuberculosis* Erdman cDNA expression library described above. Five reactive clones, referred to as Tb431, Tb472, Tb470, Tb838 and Tb962 were isolated. The determined cDNA sequences for Tb431, Tb472, Tb470, and Tb838 are provided in SEQ ID NO: 11, 12, 104 and 105, respectively, with the determined cDNA sequences for Tb962 being provided in SEQ ID NO: 106 and 107. The corresponding predicted amino acid sequence for Tb431 is provided in SEQ ID NO: 15.

Subsequent studies led to the isolation of a full-length cDNA sequence for Tb472 (SEQ ID NO: 108). Overlapping peptides were synthesized and used to identify the reactive open reading frame. The predicted amino acid sequence for the protein encoded by Tb472 (referred to as MSL) is provided in SEQ ID NO: 109. Comparison of the sequences for Tb472 and MSL with those in the gene bank, as described above, revealed no homologies to known sequences. Fifteen overlapping peptides to the sequence of MSL (referred to as MSL-1 – MSL-15; SEQ ID NO: 110-124, respectively) were synthesized and their ability to stimulate T cell proliferation and IFN-γ production in a CD4+ T cell line generated against *M. tuberculosis* culture filtrate was examined as described below. The peptides referred to as MSL-10 (SEQ ID NO: 119) and MSL-11 (SEQ ID NO: 120) were found to show the highest level of reactivity.

Comparison of the determined cDNA sequence for Tb838 with those in the gene bank revealed identity to the previously isolated *M. tuberculosis* cosmid SCY07H7. Comparison of the determined cDNA sequences for the clone Tb962 with those in the gene

bank revealed some homology to two previously identified *M. tuberculosis* cosmids, one encoding a portion of bacteoferritin. However, recombinant bacteoferritin was not found to be reactive with the T cell line used to isolate Tb962.

The clone Tb470, described above, was used to recover a full-length open reading (SEQ ID NO: 125) that showed homology with TbH9 and was found to encode a 40 kDa antigen, referred to as Mtb40. The determined amino acid sequence for Mtb40 is provided in SEQ ID NO: 126. Similarly, subsequent studies led to the isolation of the full-length cDNA sequence for Tb431, provided in SEQ ID NO: 83, which was also determined to contain an open reading frame encoding Mtb40. Tb470 and Tb431 were also found to contain a potential open reading frame encoding a U-ORF-like antigen.

Screening of an *M. tuberculosis* Erdman cDNA expression library with multiple CD4+ T cell lines generated against *M. tuberculosis* culture filtrate, resulted in the isolation of three clones, referred to as Tb366, Tb433 and Tb439. The determined cDNA sequences for Tb366, Tb433 and Tb439 are provided in SEQ ID NO: 127, 128 and 129, respectively. Comparison of these sequences with those in the gene bank revealed no significant homologies to Tb366. Tb433 was found to show some homology to the previously identified *M. tuberculosis* antigen MPT83. Tb439 was found to show 100% identity to the previously isolated *M. tuberculosis* cosmid SCY02B10.

A CD4+ T cell line was generated against *M. tuberculosis* PPD, essentially described above, and used to screen the above *M. tuberculosis* Erdman cDNA expression library. One reactive clone (referred to as Tb372) was isolated, with the determined cDNA sequences being provided in SEQ ID NO: 130 and 131. Comparison of these sequences with those in the gene bank revealed no significnat homologies.

In further studies, screening of an *M. tuberculosis* cDNA expression library with a CD4+ T cell line generated against dendritic cells that had been infected with tuberculosis for 8 days, as described above, led to the isolation of two clones referred to as Tb390R5C6 and Tb390R2C11. The determined cDNA sequence for Tb390R5C6 is provided in SEQ ID NO: 132, with the determined cDNA sequences for Tb390R2C11 being provided in SEQ ID NO: 133 and 134. Tb390R5C6 was found to show 100% identity to a previously identified *M. tuberculosis* cosmid.

In subsequent studies, the methodology described above was used to screen an *M. tuberculosis* genomic DNA library prepared as follows. Genomic DNA from *M. tuberculosis* Erdman strain was randomly sheared to an average size of 2 kb, and blunt ended with Klenow polymerase, followed by the addition of EcoRI adaptors. The insert was subsequently ligated into the Screen phage vector (Novagen, Madison, WI) and packaged *in vitro* using the PhageMaker extract (Novagen). The phage library (referred to as the Erd λ Screen library) was amplified and a portion was converted into a plasmid expression library by an autosubcloning mechanism using the *E. coli* strain BM25.8 (Novagen). Plasmid DNA was purified from BM25.8 cultures containing the pSCREEN recombinants and used to transform competent cells of the expressing host strain BL21(DE3)pLysS. Transformed cells were aliquoted into 96 well microtiter plates with each well containing a pool size of approximately 50 colonies. Replica plates of the 96 well plasmid library format were induced with IPTG to allow recombinant protein expression. Following induction, the plates were centrifuged to pellet the *E. coli* which was used directly in T cell expression cloning of a CD4+ T cell line prepared from a PPD-positive donor (donor 160) as described above. Pools containing *E. coli* expressing *M. tuberculosis* T cell antigens were subsequently broken down into individual colonies and reassayed in a similar fashion to identify positive hits.

Screening of the T cell line from donor 160 with one 96 well plate of the Erd λ Screen library provided a total of nine positive hits. Previous experiments on the screening of the pBSK library described above with T cells from donor 160 suggested that most or all of the positive clones would be TbH-9, Tb38-1 or MTI (disclosed in U.S. Patent Application No. 08/533,634) or variants thereof. However, Southern analysis revealed that only three wells hybridized with a mixed probe of TbH-9, Tb38-1 and MTI. Of the remaining six positive wells, two were found to be identical. The determined 5' cDNA sequences for two of the isolated clones (referred to as Y1-26C1 and Y1-86C11) are provided in SEQ ID NO: 135 and 136, respectively. The full length cDNA sequence for the isolated clone referred to as hTec#1 is provided in SEQ ID NO: 137, with the corresponding predicted amino acid sequence being provided in SEQ ID NO: 138. Comparison of the sequences of hTec#1 to those in the gene bank as described above, revealed some homology to the previously isolated *M. tuberculosis* cosmid MTCY07H7B.06

EXAMPLE 2

INDUCTION OF T CELL PROLIFERATION AND INTERFERON-γ PRODUCTION BY *M. TUBERCULOSIS* ANTIGENS

The ability of recombinant *M. tuberculosis* antigens to induce T cell proliferation and interferon-γ production may be determined as follows.

Proteins may be induced by IPTG and purified by gel elution, as described in Skeiky et al. *J. Exp. Med.*, 1995, *181*:1527-1537. The purified polypeptides are then screened for the ability to induce T-cell proliferation in PBMC preparations. The PBMCs from donors known to be PPD skin test positive and whose T-cells are known to proliferate in response to PPD, are cultured in medium comprising RPMI 1640 supplemented with 10% pooled human serum and 50 μg/ml gentamicin. Purified polypeptides are added in duplicate at concentrations of 0.5 to 10 μg/mL. After six days of culture in 96-well round-bottom plates in a volume of 200 μl, 50 μl of medium is removed from each well for determination of IFN-γ levels, as described below. The plates are then pulsed with 1 μCi/well of tritiated thymidine for a further 18 hours, harvested and tritium uptake determined using a gas scintillation counter. Fractions that result in proliferation in both replicates three fold greater than the proliferation observed in cells cultured in medium alone are considered positive.

IFN-γ is measured using an enzyme-linked immunosorbent assay (ELISA). ELISA plates are coated with a mouse monoclonal antibody directed to human IFN-γ (PharMingen, San Diego, CA) in PBS for four hours at room temperature. Wells are then blocked with PBS containing 5% (W/V) non-fat dried milk for 1 hour at room temperature. The plates are washed six times in PBS/0.2% TWEEN-20 and samples diluted 1:2 in culture medium in the ELISA plates are incubated overnight at room temperature. The plates are again washed and a polyclonal rabbit anti-human IFN-γ serum diluted 1:3000 in PBS/10% normal goat serum is added to each well. The plates are then incubated for two hours at room temperature, washed and horseradish peroxidase-coupled anti-rabbit IgG (Sigma Chemical So., St. Louis, MO) is added at a 1:2000 dilution in PBS/5% non-fat dried milk. After a further two hour incubation at room temperature, the plates are washed and TMB substrate

added. The reaction is stopped after 20 min with 1 N sulfuric acid. Optical density is determined at 450 nm using 570 nm as a reference wavelength. Fractions that result in both replicates giving an OD two fold greater than the mean OD from cells cultured in medium alone, plus 3 standard deviations, are considered positive.

EXAMPLE 3

PURIFICATION AND CHARACTERIZATION OF *M. TUBERCULOSIS* POLYPEPTIDES USING CD4+ T CELL LINES GENERATED FROM A MOUSE *M. TUBERCULOSIS* MODEL

Infection of C57BL/6 mice with *M. tuberculosis* results in the development of a progressive disease for approximately 2-3 weeks. The disease progression is then halted as a consequence of the emergence of a strong protective T cell-mediated immune response. This infection model was used to generate T cell lines capable of recognizing protective *M. tuberculosis* antigens.

Specifically, spleen cells were obtained from C57BL/6 mice infected with *M. tuberculosis* for 28 days and used to raise specific anti-*M. tuberculosis* T cell lines as described above. The resulting CD4+ T cell lines, in conjunction with normal antigen presenting (spleen) cells from C57BL/6 mice were used to screen the *M. tuberculosis* Erd λ screen library described above. One of the reactive library pools, which was found to be highly stimulatory of the T cells, was selected and the corresponding active clone (referred to as Y288C10) was isolated.

Sequencing of the clone Y288C10 revealed that it contains two potential genes, in tandem. The determined cDNA sequences for these two genes (referred to as mTCC#1 and mTCC#2) are provided in SEQ ID NO: 139 and 140, respectively, with the corresponding predicted amino acid sequences being provided in SEQ ID NO: 141 and 142, respectively. Comparison of these sequences with those in the gene bank revealed identity to unknown sequences previously found within the *M. tuberculosis* cosmid MTY21C12. The predicted amino acid sequences of mTCC#1 and mTCC#2 were found to show some homology to previously identified members of the Tb38 protein family, discussed above.

EXAMPLE 4

SYNTHESIS OF SYNTHETIC POLYPEPTIDES

Polypeptides may be synthesized on a Millipore 9050 peptide synthesizer using FMOC chemistry with HPTU (O-Benzotriazole-N,N,N',N'-tetramethyluronium hexafluorophosphate) activation. A Gly-Cys-Gly sequence may be attached to the amino terminus of the peptide to provide a method of conjugation or labeling of the peptide. Cleavage of the peptides from the solid support may be carried out using the following cleavage mixture: trifluoroacetic acid:ethanedithiol:thioanisole:water:phenol (40:1:2:2:3). After cleaving for 2 hours, the peptides may be precipitated in cold methyl-t-butyl-ether. The peptide pellets may then be dissolved in water containing 0.1% trifluoroacetic acid (TFA) and lyophilized prior to purification by C18 reverse phase HPLC. A gradient of 0-60% acetonitrile (containing 0.1% TFA) in water (containing 0.1% TFA) may be used to elute the peptides. Following lyophilization of the pure fractions, the peptides may be characterized using electrospray mass spectrometry and by amino acid analysis.

EXAMPLE 4

USE OF REPRESENTATIVE ANTIGENS FOR SERODIAGNOSIS OF TUBERCULOSIS

The diagnostic properties of representative *M. tuberculosis* antigens may be determined by examining the reactivity of antigens with sera from tuberculosis-infected patients and from normal donors as described below.

Assays are performed in 96-well plates coated with 200 ng antigen diluted to 50 µL in carbonate coating buffer, pH 9.6. The wells are coated overnight at 4°C (or 2 hours at 37°C). The plate contents are then removed and the wells are blocked for 2 hours with 200 µL of PBS/1% BSA. After the blocking step, the wells are washed five times with PBS/0.1% Tween 20™. 50 µL sera, diluted 1:100 in PBS/0.1% Tween 20™/0.1% BSA, is

then added to each well and incubated for 30 minutes at room temperature. The plates are washed again five times with PBS/0.1% Tween 20™.

The enzyme conjugate (horseradish peroxidase - Protein A, Zymed, San Francisco, CA) is then 1:10,000 in PBS/0.1% Tween 20™/0.1% BSA, and 50 µL of the diluted conjugate is added to each well and incubated for 30 minutes at room temperature. Following incubation, the wells are washed five times with PBS/0.1% Tween 20™. 100 µL of tetramethylbenzidine peroxidase (TMB) substrate (Kirkegaard and Perry Laboratories, Gaithersburg, MD) is added, undiluted, and incubated for about 15 minutes. The reaction is stopped with the addition of 100 µL of 1 N H_2SO_4 to each well, and the plates are read at 450 nm.

From the foregoing, it will be appreciated that, although specific embodiments of the invention have been described herein for the purpose of illustration, various modifications may be made without deviating from the spirit and scope of the invention.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANTS: Alderson, Mark
Dillon, Davin C.
Skeiky, Yasir A.W.
Campos-Neto, Antonio

(ii) TITLE OF INVENTION: COMPOUNDS AND DIAGNOSIS OF TUBERCULOSIS AND METHODS OF THEIR USE

(iii) NUMBER OF SEQUENCES: 144

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: SEED and BERRY
(B) STREET: 6300 Columbia Center, 701 Fifth Ave.
(C) CITY: Seattle
(D) STATE: Washington
(E) COUNTRY: US
(F) ZIP: 98104

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE: 08-MAY-1998
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Maki, David J.
(B) REGISTRATION NUMBER: 31,392
(C) REFERENCE/DOCKET NUMBER: 210121.441C1

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 206-622-4900
(B) TELEFAX: 206-682-6031

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1886 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
CGCTCTGGTG ACCACCAACT TCTTCGGTGT CAACACCATC CCGATCGCCC TCAACGAGGC      60
CGACTACCTG CGCATGTGGA TCCAGCCCGC CACCGTCATG AGCCACTATC AAGCCGTCGC     120
GCACGAAATC TGGTGTCTCC ATGAATATGC CAGTTCCGGA AAGCCGTGGG CCAGTATCAC     180
CACGGGTGCG CCGGGCTCAC CGGCCTCGAC CACTCGCAGT CGCACGCCGT TGGTATCAAC     240
TAACCGTNCN GTANGTGCGC CCATCGTCTC ACCAAATCAC ACCGGGCACC GGCCTGAGAA     300
GGGCTTGGGG AGCANCCAGA GGCGATTGTC GCGGGTGCTG CCGCGCATCA TTGATCGGCC     360
GGCCGGACCA NTCGGGCCTC CCTTGACGTC CGGATCNCAC TTCCTGTGCA GCTGGCATGG     420
CTACAGCTCA CAGTGACTGC CCCACGATTG CCGGCCAGGT CCAGTTCAAA TTCCGGTGAA     480
TTCGCCGACA AAAGCAGCAG GTCAACCAAC CGCAGTCAGT CGAGGGTCCC AAACGTGAGC     540
CAATCGGTGA AATGGCTTGC TGCAGTGACA CCGGTCACAG GCTTAGCCGA CAGCACCGGA     600
ATAGCTCAGG CGGGCTATAG AGTCCTATAG AAACATTTGC TGATAGAATT AACCGCTGTC     660
TTGGCGTGAT CTTGATACGG CTCGCCGTGC GACCGGTTGG CTCAGTAGCT GACCACCATG     720
TAACCCATCC TCGGCAGGTG TCTACTAAGG CGAGACACCG CATTGGTGGG GCTGCATCGC     780
AAATCGGTCC GAGCATGTAG CACTGCCGTT ATCCCGGGAT AGCAAACCAC CCGGAACCAG     840
GGCTATCCCA GTCGCTCTCC GACGGAGGCC GTTTCGCTTT CCGTTGCCCG ATAACTCCCG     900
AGTGGATATC GGCGTTATCA NATTCAGGCT TTTCTTCGCA AGGTACCGGT GTTCGCTATA     960
TTCGGATATC TCGGACGGAT AATTACTAAA ACTTCAGTGG TTTAGATAAG GCCGCCGCAA    1020
TACTTCGCCG ATCTTGCCGA GCGCAACGGA TTTCCATCGT CGGTTTTCGT CGCCTTATCA    1080
AACATGATCG GAGATAATGA CAGATCGGCC TAGCTAGGTG TTTAGCCGAC GCGATTTAGG    1140
ACAACCGAGA TTTGCTTTGC CTCGCAACCA TGAGAGCGCC CCGCTTCGAC GCCGAATCGG    1200
GTGAGTGATG GTGGGTTAGC ACAGCCCTGA TTGCGCCAGC GGCGAGGTGA TTGTGCCCGC    1260
CACGAGGCCG CCGCCGGCTA GCCCCATGAG CACGNTATAT AGACTCTCCT GCAACAGATC    1320
TCATACCGAT CGAAGGCGAA GCGCAGGCAT CGACGTCGGA GACACTGCCT TGGGATCGCG    1380
CCGCCTACAC GGCGGTTGGC GCATTGTCGC AGCGCAGTTG CAGGAGGGCA AATGTYGCGA    1440
GACGATGTAG TCGACAACAA GTGNACATGC CGTCTTCACG AACTCAAAAC TGACGATCTG    1500
CTTAGCATGA AAAAAACTGT TGACATCGGC CAAGCATGAC AGCCAGACTG TAGGCCTACG    1560
CGTGCAATGC AGAACCAAGG NTATGCATGG AATCGACGAC CGTTGAGATA GGCGGCAGGC    1620
ATGAGCAGAG CGTTCATCAT CGATCCAACG ATCAGTGCCA TTGACGGCTT GTACGACCTT    1680
CTGGGGATTG GAATACCCAA CCAAGGGGGT ATCCTTTACT CCTCACTAGA GTACTTCGAA    1740
AAAGCCCTGG AGGAGCTGGC AGCAGCGTTT CCGGGTGATG GCTGGTTAGG TTCGGCCGCG    1800
GACAAATACG CCGGCAAAAA CCGCAACCAC GTGAATTTTT TCCAGGAACT GGCAGACCTC    1860
GATCGTCAGC TCATCAGCCT GATCCA                                         1886
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2305 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
GGCACGCGCT GGCCGCGCAA TACACCGAAA TTGCAACGGA ACTCGCAAGC GTGCTCGCTG      60
CGGTGCAGGC AAGCTCGTGG CAGGGGCCCA GCGCCGACCG GTTCGTCGTC GCCCATCAAC     120
CGTTCCCGGTA TTGGCTAACC CACGCTGCCA CGGTGGCCAC CGCAGCAGCC GCCGCGCACN     180
```

```
AAACGGCCGC CGCCGGGTAT ACGTCCGCAT TGGGGGCAT GCCTACGCTA GCCGAGTTGG      240
CGGCCAACCA TGCCATGCAC GGCGCTCTGG TGACCACCAA CTTCTTCGGT GTCAACACCA      300
TCCCGATCGC CCTCAACGAG GCCGACTACC TGCGCATGTG GATCCAGGCC GCCACCTTCA      360
TGAGCCACTA TCAAGCCGTC GCGCACGAAA GCGTGGCGGC GACCCCCAGC ACGCCGCCGG      420
CGCCGCAGAT AGTGACCAGT GCGGCCAGCT CGGCGGCTAG CAGCAGCTTC CCCGACCCGA      480
CCAAATTGAT CCTGCAGCTA CTCAAGGATT TCCTGGAGCT GCTGCGCTAT CTGGCTGTTG      540
AGCTGCTGCC GGGGCCGCTC GGCGACCTCA TCGCCCAGGT GTTGGACTGG TTCATCTCGT      600
TCGTGTCCGG TCCAGTCTTC ACGTTTCTCG CCTACCTGGT GCTGGACCCA CTGATCTATT      660
TCGGACCGTT CGCCCCGCTG ACGAGTCCGG TCTTGTTGCC TGCTGTGGAG TTACGCAACC      720
GCCTCAAAAC CGCCACCGGA CTGACGCTGC CACCTACCGT GATTTTCGAT CATCCCACTC      780
CCACTGCGGT CGCCGAGTAT GTCGCCCAGC AAATGTCTGG CAGCCGCCCA ACGGAATCCG      840
GTGATCCGAC GTCGCAGGTT GTCGAACCCG CTCGTGCCGA ATTCGGCACG AGTGCTGTTC      900
ATCAAATCCC CCCGAGACCT GCGGACACCC GGCGCGCTTG CCGACATCGA GATGATGTCC      960
CGCGAGATAG CAGAATTGCC CAACATCGTG ATGGTGCGGG GCTTGACCCG ACCGAACGGG     1020
GAACCTCTGA AGGAGACCAA GGTCTCGTTT CAGGCTGGTG AAGTGGGCGG CAAGCTCGAC     1080
GAAGCGACCA CCCTGCTCGA AGAGCACGGA GGCGAGCTGG ACCAGCTGAC CGGCGGTGCG     1140
CACCAGTTGG CCGACGCCCT CGCCCAAATA CGCAACGAAA TCAATGGGGC CGTGGCCAGC     1200
TCGAGCGGGA TAGTCAACAC CCTGCAGGCC ATGATGGACC TGATGGGCGG TGACAAGACC     1260
ATCCGACAAC TGGAAAATGC GTCCCAATAT GTCGGGCGCA TGCCGGCTCT GGGGGACAAT     1320
CTGAGCGGGA CCGTCACCGA TGCCGAACAA ATCGCCACTT GGGCCAGCCC TATGGTCAAC     1380
GCCCTCAACT CCAGCCCGGT GTGTAACAGC GATCCCGCCT GTCGGACGTC GCGCGCACAG     1440
TTGGCGGCGA TTGTCCAGGC GCAGGACGAC GGCCTGCTCA GGTCCATCAG AGCGCTAGCC     1500
GTCACCCTGC AACAGACGCA GGAATACCAG ACACTCGCCC GGACGGTGAG CACACTGGAC     1560
GGGCAACTGA AGCAAGTCGT CAGCACCCTC AAAGCGGTCG ACGGCCTACC CACCAAATTG     1620
GCTCAAATGC AGCAAGGAGC CAACGCTCTC GCCGACGGCA GCGCAGCGCT GGCGGCAGGC     1680
GTGCAGGAAT TGGTCGATCA GGTCAAAAAG ATGGGCTCAG GGCTCAACGA GGCCGCCGAC     1740
TTCCTGTTGG GGATCAAGCG GGATGCGGAC AAGCCGTCAA TGGCGGGCTT CAACATTCCA     1800
CCGCAGATTT TTTCGAGGGA CGAGTTCAAG AAGGGCGCCC AGATTTTCCT GTCGGCCGAT     1860
GGTCATGCGG CGCGGTACTT CGTGCAGAGC GCGCTGAATC CGGCCACCAC CGAGGCGATG     1920
GATCAGGTCA ACGATATCCT CCGTGTTGCG GATTCCGCGC GACCGAATAC CGAACTCGAG     1980
GATGCCACGA TAGGTCTGGC GGGGGTTCCG ACTGCGCTGC GGGATATCCG CGACTACTAC     2040
AACAGCGATA TGAAATTCAT CGTCATTGCG ACGATCGTTA TCGTATTCTT GATTCTCGTC     2100
ATTCTGNTGC GCGCACTTGT CGNTCCGATA TATCTGATAG GCTCGGTGCT GATTTCTTAC     2160
TTGTCGGCCC TAGGCATAGG AACTTTCGTT TTCCAATTGA TACTGKGCCA GGAAATGCAT     2220
TGGAGCCTGC CGGGACTGTC CTTCATATTA TTGGTTGCCA TCGGCGCTGA CTACAACATG     2280
CTGCTCATTT CACGCATCCG CGACG                                           2305
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1742 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
CCGCTCTCTT TCAACGTCAT AAGTTCGGTG GGCCAGTCGG CCGCGCGTGC ATATGGCACC       60
AATAACGCGT GTCCCATGGA TACCCGGACC GCACGACGGT AGAGCGGATC AGCGCAGCCG      120
GTGCCGAACA CTACCGCGTC CACGCTCAGC CCTGCCGCGT TGCCGGAAGAT CGAGCCCAGG     180
```

```
TTCTCATGGT CGTTAACGCC TTCCAACACT GCGACGGTGC GCGCCCCGGC GACCACCTGA      240
GCAACGCTCG GCTCCGGCAC CCGGCCGCGG GCTGCCAACA CCCCACGATT GAGATGGAAG      300
CCGATCACCC GTGCCATGAC ATCAGCCGAC GCTCGATAGT ACGGCCGCCC GACACCGGCC      360
AGATCATCCT TGAGCTCGGC CAGCCGGCGG TCGGTGCCGA ACAGCGCCAG CGGCGTGAAC      420
CGTGAGGCCA GCATGCGCTG CACCACCAGC ACACCCTCGG CGATCACCAA CGCCTTGCCG      480
GTCGGCAGAT CGGGACNACN GTCGATGCTG TTCAGGTCAC GGAAATCGTC GAGCCGTGGG      540
TCGTCGGGAT CGCAGACGTC CTGAACATCG AGGCCGTCGG GGTGCTGGGC ACAACGGCCT      600
TCGGTCACGG GCTTTCGTCG ACCAGAGCCA GCATCAGATC GGCGGCGCTG CGCAGGATGT      660
CACGCTCGCT GCGGTTCAGC GTCGCGAGCC GCTCAGCCAG CCACTCTTGC ACAGAGCCGT      720
TGCTGGGATT AATTGGGAGA GGAAGACAGC ATGTCGTTCG TGACCACACA GCCGGAAGCC      780
CTGGCAGCTG CGGCGGCGAA CCTACAGGGT ATTGGCACGA CAATGAACGC CCAGAACGCG      840
GCCGCGGCTG CTCCAACCAC CGGAGTAGTG CCCGCAGCCG CCGATGAAGT ATCAGCGCTG      900
ACCGCGGCTC AGTTTGCTGC GCACGCGCAG ATGTACCAAA CGGTCAGCGC CCAGGCCGCG      960
GCCATTCACG AAATGTTCGT GAACACCCTG GTGGCCAGTT CTGGCTCATA CGCGGCCACC     1020
GACGCGGCCA ACGCAGCCGC TGCCGGCTGA ACGGGCTCGC ACGAACCTGC TGAAGGAGAG     1080
GGCGAACATC CGGAGTTCTC GGGTCAGGGG TTGCGCCAGC GCCCAGCCGA TTCAGNTATC     1140
GGCGTCCATA ACAGCAGACG ATCTAGGCAT TCAGTACTAA GGAGACAGGC AACATGGCCT     1200
CACGTTTTAT GACGGATCCG CATGCGATGC GGGACATGGC GGGCCGTTTT GAGGTGCACG     1260
CCCAGACGGT GGAGGACGAG GCTCGCCGGA TGTGGCGTTC CGCGCAAAAC ATTTCCGGTG     1320
CGGGCTGGAG TGGCATGGCC GAGGCGACCT CGCTAGACAC CATGACCTAG ATGAATCAGG     1380
CGTTTCGCAA CATCGTGAAC ATGCTGCACG GGGTGCGTGA CGGGCTGGTT CGCGACGCCA     1440
ACAANTACGA ACAGCAAGAG CAGGCCTCCC AGCAGATCCT GAGCAGNTAG CGCCGAAAGC     1500
CACAGCTGNG TACGNTTTCT CACATTAGGA GAACACCAAT ATGACGATTA ATTACCAGTT     1560
CGGGGACGTC GACGCTCATG GCGCCATGAT CCGCGCTCAG GCGGCGTCGC TTGAGGCGGA     1620
GCATCAGGCC ATCGTTCGTG ATGTGTTGGC CGCGGGTGAC TTTTGGGGCG GCGCCGGTTC     1680
GGTGGCTTGC CAGGAGTTCA TTACCCAGTT GGGCCGTAAC TTCCAGGTGA TCTACGAGCA     1740
GG                                                                    1742
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2836 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
GTTGATTCCG TTCGCGGCGC CGCCGAAGAC CACCAACTCC GCTGCGGTGG TCGCACAGGC       60
GGTTGCGTCG GTCAGCTGGC CGAATCCCAA TGATTGGTGG CTCNGTGCGG TTGCTGGCCT      120
CGATTACCCC CACGGAAAGG ACGACGATCC TTCGTTTGCT CGGTCAGTCG TACTTGGCGA      180
CGGGCATGGC GCGGTTTCTT ACCTCGATCG CACAGCAGCT GACCTTCGGC CCAGGGGGCA      240
CAACGGCTGG CTCCGGCGGA GCCTGGTACC CAACGCCACA ATTCGCCGGC CTGGTTGCAG      300
GCCCGGCCGT GTCGGCGAGT TTGGCGCGGG CGGAGCCGGT CGGGAGGTTG TCGGTGCCGC      360
CAAGTTGGGC CGTCGCGGCT CCGGCCTTCG CGCAGAAGCC TGAGGCCGGC ACGCCGATGT      420
CCGTCATCGG CGAAGCGTCC AGCTGCGGTC AGGGAGGCCT GCTTCGAGGC ATACCGCTGG      480
CGAGAGCGGG GCGGCGTACA GGCGCCTTCG CTCACCGATA CGGGTTCCGC CACAGCGTGA      540
TTACCCGGTC TCCGTCGGCG GGATAGCTTT CGATCCGGTC TGCGCGGCCG CCGGAAATGC      600
TGCAGATAGC GATCGACCGC GCCGGTCGGT AAACGCCGCA CACGGCACTA TCAATGCGCA      660
CGGCGGGCGT TGATGCCAAA TTGACCGTCC CGACGGGGCT TTATCTGCGG CAAGATTTCA      720
```

```
TCCCCAGCCC GGTCGGTGGG CCGATAAATA CGCTGGTCAG CGCGACTCTT CCGGCTGAAT      780
TCGATGCTCT GGGCGCCCGC TCGACGCCGA GTATCTCGAG TGGGCCGCAA ACCCGGTCAA      840
ACGCTGTTAC TGTGGCGTTA CCACAGGTGA ATTTGCGGTG CCAACTGGTG AACACTTGCG      900
AACGGGTGGC ATCGAAATCA ACTTGTTGCG TTGCAGTGAT CTACTCTCTT GCAGAGAGCC      960
GTTGCTGGGA TTAATTGGGA GAGGAAGACA GCATGTCGTT CGTGACCACA CAGCCGGAAG     1020
CCCTGGCAGC TGCGGCGGCG AACCTACAGG GTATTGGCAC GACAATGAAC GCCCAGAACG     1080
CGGCCGCGGC TGCTCCAACC ACCGGAGTAG TGCCCGCAGC CGCCGATGAA GTATCAGCGC     1140
TGACCGCGGC TCAGTTTGCT GCGCACGCGC AGATGTACCA AACGGTCAGC GCCCAGGCCG     1200
CGGCCATTCA CGAAATGTTC GTGAACACGC TGGTCGCCAG TTCTGGCTCA TACGCGGCCA     1260
CCGAGGCGGC CAACGCAGCC GCTGCCGGCT GAACCGGCTC GCACGAACCT GCTGAAGGAG     1320
AGGCGGAACA TCCGGAGTTC TCGGGTCAGG GGTTGCGCCA GCGCCCAGCC GATTCAGCTA     1380
TCGCCGTCCA TAACAGCAGA CGATCTAGGC ATTCAGTACT AAGGAGACAG GCAACATGGC     1440
CTCACGTTTT ATGACGGATC CGCATGCGAT GCCGGACATG GCGGGCCGTT TTGAGGTGCA     1500
CGCCCAGACG GTGGAGGACG AGGCTCGCCG GATGTGGGCG TCCGCGCAAA ACATTTCCGG     1560
TGCGGGCTGG AGTGGCATGG CCGAGGCGAC CTCGCTAGAC ACCATGACCT AGATGAATCA     1620
GGCGTTTCGC AACATCGTGA ACATGCTGCA CGGGGTGCGT GACGGGCTGG TTCGCGACGC     1680
CAACAACTAC GAACAGCAAG AGCAGCGCTC CCAGCAGATC CTGAGCAGCT AGCGCCGAAA     1740
GCCACAGCTG CGTACGCTTT CTCACATTAG GAGAACACCA ATATGACGAT TAATTACCAG     1800
TTCGGGGACG TCGACGCTCA TGGCGCCATG ATCCGCGCTC AGGCGGCGTC GCTTGAGGCG     1860
GAGCATCAGG CCATCGTTCG TGATGTGTTG GCCGCGGGTG ACTTTTCGGG CGGCGCCGGT     1920
TCGGTTGCTT GCCAGGAGTT CATTACCCAG TTGGGCCGTA ACTTCCAGGT GATCTACGAG     1980
CAGGCCAACG CCCACGGGCA GAAGGTGCAG GCTGCCGGCA ACAACATGGC GCAAACCGAC     2040
AGCGCCGTCG GCTCCAGCTG GGCCTAAAAC TGAACTTCAG TCGCGGCAGC ACACCAACCA     2100
GCCGGTGTGC TGCTGTGTCC TGCAGTTAAC TAGCACTCGA CCGCTGACGT AGCGATCGAT     2160
CAACAGAGTA CCCGCACCGA CATCACCGTC AACGTCGACG GCTTCTGGAT GCTTCAGGCG     2220
CTACTGGATA TCCGCCACGT TGCGCCTGAG TTACGTTGCC GGCCGTACGT CTCCACCGAT     2280
TCCAATGACT GGCTAAACGA GCACCCGCGG ATGGCGGTCA TGCGCGAGCA GGGCATTGTC     2340
GTCAACGACG CGGTCAACGA ACAGGTCGCT GCCCCGATGA AGGTGCTTGC CGCACCTGAT     2400
CTTGAAGTCG TCGCCCTGCT GTCACGCGGC AAGTTGCTGT ACGGGGTCAT AGACGACGAG     2460
AACCAGCCGC CGGGTTCGCG TGACATCCCT GACAATGAGT TCCGGGTGGT GTTGGCCCGG     2520
CGAGGCCAGC ACTGGGTGTC GCCGGTACGG GTTGGCAATG ACATCACCGT CGATGACGTG     2580
ACGGTCTCGG ATAGCGCCTC GATCGCCGCA CTGGTAATGG ACGGTCTGGA GTCGATTCAC     2640
CACGCCGACC CAGCCGCGAT CAACGCGGTC AACGTGCCAA TGGAGGAGAT CTCGTGCCGA     2700
ATTCGGCACG AGGCACGAGG CGGTGTCGGT GACGACGGGA TCGATCACGA TCATCGACCG     2760
GCCGGGATCC TTGGCGATCT CGTTGAGCAC GACCCGGGCC CGCGGGAAGC TCTGCGACAT     2820
CCATGGGTTC TTCCCG                                                     2836
```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 900 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
AACATGCTGC ACGGGGTGCG TGACGGGCTG GTTCGCGACG CCAACAACTA CGAGCAGCAA       60
GAGCAGGCCT CCCAGCAGAT CCTCAGCAGC TAACGTCAGC CGCTGCAGCA CAATACTTTT      120
ACAAGCGAAG GAGAACAGGT TCGATGACCA TCAACTATCA GTTCGGTGAT GTCGACGCTC      180
```

```
ACGGCGCCAT GATCCGCGCT CAGGCCGGGT TGCTGGAGGC CGAACATCAG GCCATCATTC      240
GTGATGTGTT GACCGCGACT GACTTTTGGG GCGGCGCCGG TTCGGCGGCC TGCCAGGGGT      300
TCATTACCCA ATTGGGCCGT AACTTCCAGG TGATCTACGA ACAGGCCAAC GCCCACGGGC      360
AGAAGGTGCA GGCTGCCGGC AACAACATGG CGCAAACCGA CAGCGCCGTC GGCTCCAGCT      420
GGGCCTGACA CCAGGCCAAG GCCAGGGACG TGGTGTACGA GTGAAGGTTC CTCGCGTGAT      480
CCTTCGGGTG GCAGTCTAGG TNGTCAGTGC TGGGGTGTTG GTGGTTTGCT GCTTGGCGGG      540
TTCTTCGGTG CTGGTCAGTG CTGCTCGGGC TCGGGTGAGG ACCTCGAAGC CCAGGTAGCG      600
CCGTCCTTCG ATCCATTCGT CGTGTTGTTC GGCGAGGACG GCTCCGACGA GGCGGATGAT      660
CGAGGCGCGG TCGGGGAAGA TGCCCACGAC GTCGGTTCGG CGTCGTACCT CTCGGTTGAG      720
GCGTTCCTGG GGGTTGTTGG ACCAGATTTG GCGCCAGATC TTCTTGGGGA AGGCGGTGAA      780
CGCCAGCAGG TCGGTGCGGG CGGTGTCGAN GTGCTCGGCC ACCGCGGGGA GTTTGTCGGT      840
CAGAGCGTCG AGTACCCGAT CATATTGGGC AACAACTGAT TCGGCGTTGG GCTGGTCGTA      900

(2) INFORMATION FOR SEQ ID NO:6:

     (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 1985 base pairs
          (B) TYPE: nucleic acid
          (C) STRANDEDNESS: double
          (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

    (vi) ORIGINAL SOURCE:
          (A) ORGANISM: Mycobacterium tuberculosis

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GCTCGCCGGA TGTGGGCGTC CGCGCAAAAC ATTTCCGGTG CGGGCTGGAG TGGCATGGCC       60
GAGGCGACCT CGCTAGACAC CATGGCCCAG ATGAATCAGG CGTTTCGCAA CATCGTGAAC      120
ATGCTGCACG GGGTGCGTGA CGGGCTGGTT CGCGACGCCA ACAACTACGA GCAGCAAGAG      180
CAGGCCTCCC AGCAGATCCT CAGCAGCTAA CGTCAGCCGC TGCAGCACAA TACTTTTACA      240
AGCGAAGGAG AACAGGTTCG ATGACCATCA ACTATCAGTT CGGTGATGTC GACGCTCACG      300
GCGCCATGAT CCGCGCTCAG GCCGGGTTGC TGGAGGCCGA GCATCAGGCC ATCATTCGTG      360
ATGTGTTGAC CGCGAGTGAC TTTTGGGGCG GCGCCGGTTC GGCGGCCTGC CAGGGGTTCA      420
TTACCCAGTT GGGCCGTAAC TTCCAGGTGA TCTACGAACA AGCCAACACC CACGGGCAGA      480
AGGTGCAAGC TGCCGGCAAC AACATGGCGC AAACCGACAG CGCCGTCGGC TCCAGCTGGG      540
CCTGACACCA GGCCAAGGCC AGGGACGTGG TGTACGAGTG AAGGTTCCTC GCGTGATCCT      600
TCGGGTGGCA GTCTAGGTGG TCAGTGCTGG GGTGTTGGTG GTTTGCTGCT TGGCGGGTTC      660
TTCGGTGCTG GTCAGTGCTG CTCGGGCTCG GGTGAGGACC TCGAGGCCCA GGTAGCGCCG      720
TCCTTCGATC CATTCGTCGT GTTGTTCGGC GAGGACGGCT CCGACGAGGC GGATGATCGA      780
GGCGCGGTCG GGGAAGATGC CCACGACGTC GGTTCGGCGT CGTACCTCTC GGTTGAAGCG      840
TTCCTGGGGG CCACCGCTTG GCGCCNAGGC ACTCCACGCC AATTCGTCNC ACCTAACAGC      900
GGTGGCCAAC GACTATGACT ACGACACCGT TTTTGCCAGG GCCCTCNAAA GGATCTGCGC      960
GTCCCGGCGA CACGCTTTTT GCGATAAGTA CCTCCGGCAA TTCTATGAGT GTACTGCGGN     1020
CCGCGAAAAC CGCAAGGGAG TTGGGTGTGA CGGTTNTTGC AAATGACGGG CGAATCCGGC     1080
GGCCAGCTGG CNGAATTCGC AGATTTCTTG ATCAACGTCC CGTCACGCGA CACCGGGCGA     1140
ATCCAGGAAT CTCACATCGT TTTTATTCAT GCGATCTCCG AACATGTCGA ACACGCGCTT     1200
TTCGCGCCTC GCCAATAGGA AAGCCGATCC TTACGCGGCC ATTCGAAAGA TGGTCGCGGA     1260
ACGTGCGGGA CACCAATGGT GTCTCTTCCT CGATAGAGAC GGGGTCATCA ATCGACAAGT     1320
GGTCGGCGAC TACGTACGGA ACTCGCGGCA GTTTGAATGG TTGCCCCGGG CGGCGCGGGC     1380
GTTGAAGAAG CTACGGGCAT GGGCTCCGTA CATCGTTGTC GTGACAAACC AGCAGGGCGT     1440
GGGTCCCGGA TTGATGAGCG CCGTCGACGT GATGGTGATA CATCGGCACC TCGAAATGCA     1500
GCTTGCATCC GATGGCGTGC TGATAGATGG ATTTCAGGTT TGCCCGCACC ACCGTTCGCA     1560
GCGGTGTGGC TGCCGTAAGC CGAGACCGGG TCTCGTCCTC GACTGGCTCG GACGACACCC     1620
```

```
CGACAGTGAG CCATTGCTGA GCATCGTGGT TGGGGACAGC CTCAGCGATC TTGACATTGG     1680
CACACAACGT CGCCGCTGCT GCCGGTGCAT GTGCCAGTGT CCAGATAGKG GGCGCCAGTT     1740
CTGGCCGGTGT CGCTGACGCG TCATTTGACT CGCTCTGGGA GTTCGCTGTC GCAGTCGGAC     1800
ATGCGCGGGG GGAGCCGGGC TAATGGCGAT CTTGCGCCGG CGACGCCGGT MGCCGGTCGG     1860
ACTNNGCGGT GGCCGGGACAG ACGTGGAACC GTACTCGAGC CAGTT                     1905

(2) INFORMATION FOR SEQ ID NO:7:

     (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 2921 base pairs
          (B) TYPE: nucleic acid
          (C) STRANDEDNESS: double
          (D) TOPOLOGY: linear

     (ii) MOLECULE TYPE: DNA (genomic)

     (vi) ORIGINAL SOURCE:
          (A) ORGANISM: Mycobacterium tuberculosis

     (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CGGGATGCCG TGGTGGTTGG TATTGCCCAA ACCCTGGCGC TGGTCCCCGG GGTATCCAGG       60
TCCGGGTCGA CCATCAGCGC TGGACTGTTT CTCGGACTCG ACCGTGAACT GGCCGCCCGA      120
TTCGGATTCC TGCTGGCCAT TCCAGCGGTG TTCGCCTCCG GGTTGTTCTC GTGGCCCGAC      180
GCATTCCACC CGGTAACCGA GGGCATGAGC GCTACTGGCC CGCAGTTGCT GGTGGCCACC      240
CTGATCGCGT TCGTCCTCGG TCTGACCGCG GTGGCCTGGC TGCTGCGGTT TCTGGTGCGA      300
CACAACATGT ACTGGTTCGT CGGCTACCGG GTGCTCGTCG GGACGGGCAT GCTCGTGCTG      360
CTGGCTACCG GGACGGTAGC CGCGACATGA CCGTCATCTT GCTACGCCAT GCCCGTTCCA      420
CCTCGAACAC CGCGGGCGTG CTGGCCGGCC GGTCCGGCGT CGACCTCGAC GAGAAGGGGC      480
GCGAGCAGGC CACCGGGTTG ATCGATCGAA TTGGTGACCT GCCGATCCGG GCGGTCGCGT      540
CTTCTCCAAT GCTGCGGTGT CAACGCACCG TCGAACCGCT GGCCGAGGCG CTGTGCCTGG      600
AGCCGCTCAT CGATGACCGG TTCTCCGAAG TCGACTACGG CGAATGGACT GGCAGAAAAA      660
TCCGTGACCT GGTCGACGAG CCGTTGTGGC GGGTAGTCCA GGCCCACCCC AGCGCGGCGG      720
TGTTTCCCGG CGGTGAGGGT TTGGCGCAGG TGCAGACGTG GTTGTCCTGA CGGATTTCCA      780
TGCCGGGGAA CACCAAGACC GGATCGGCAC TGGCGGTCGC CGGCGAAAAC CCGGCCGCCA      840
ATAGGGCGAC CGTCGCTGCG AATGCGCGTG GTACCAGGCG GACCACCTTG AACTCCCATC      900
CGTCGGGGCC AAGCGCATCG CCCGCCGCCG GTTACGGCTA AGGCGTACCA AAACCCGACG      960
GTAATACTTC GGCAATGTCG GGTCNCGACG TTACCGAGAC GTGACCAGNG AGGCNGCGGC     1020
ATTGGATTTA TCGATGGTGC GCGGTTCCCA NCCCGGCGGT CCGAANACGT AGCCCAGCCG     1080
ATCCCGCAGA CGTGTTGCCG ACCGCCAGTC ACGCACGATC GCCACGTACT CGCCGGGTCTG     1140
CAGCTTCCAG ATGTTGAACG TGTCGACCCG CTTGGTCAGG CCATAATGCG GTCCGAATAG     1200
CTCCGGCTGA AAGCTACCGA ACAGGCGGTC CCAGATGATG AGGATGCCGC CATAGTTCTT     1260
GTCCANATAC ACCGGGTCCA TTCCGTGGTG GACCCGGTGG TGCGACGGGG TATTGAAGAC     1320
GAATTCGAAC CACCGCGGCA GCCTGTCGAT CCGCTCGGTG TGCACCCAGA ACTGGTAGAT     1380
CAAGTTCAGC GACCAATTGC AGAACACCAT CCAAGGGGGA AGCCCCATCA GTGGCAGCGG     1440
AACCCACATG AGAATCTCGC CGCTGTTGTT CCANTTTCTG GCGCAGCGCG GTGGCGAAGT     1500
TGAAGTATTC GCTGGAGTGA TGCGCCTGGT GGGTAGCCCA GATCAGCCGA ACTCGGTGGG     1560
CGATGCGGTG ATAGGAGTAG TACAGCAGAT CGACACCAAC GATCGCGATC ACCCAGGTGT     1620
ACCACCGGTG GGCGGACAGC TGCCAGGCGG CAAGGTAGGC ATAGATTGCG GCATAACCGA     1680
GCAGGGCAAG GGACTTCCAG CCGGCGGTGG TGGCTATCGA AACCAGCCCC ATCGAGATGC     1740
TGGCCACCGA GTCGCGGCTG AGGTAAGCGC CCGAGGCGGG CCGTGGCTGC CCGGTAGCAG     1800
CGGTCTCGAT GCTTTCCAGC TTGCGGGCCG CCGTCCATTC GAGAATCAGC AGCAATAGAA     1860
AACATGGAAT GGCGAACAGT ACCGGGTCCC GCATTTCCTC GGGCAGCGCT GAGAAGAATC     1920
CGGCGACGGC ATGGCCGAGG CGACCTCGNT AGACACCATG ACCCAGATGA ATCAGGCGTT     1980
TCGCAACATC GTGAACATGC TGCACCGGGT GCGTGACCGG CTGGTTCGCG ACGCCAACAA     2040
```

```
NTACGAACAG CAAGAGCAGG CCTCCCAGCA GATCCTCAGC AGCTGACCCG GCCCGACGAC     2100
TCAGGAGGAC ACATGACCAT CAACTATCAA TTCGGGGACG TCGACGCTCA CGGCGCCATG     2160
ATCCGCGCTC AGGCCGGGTC GCTGGAGGCC GAGCATCAGG CCATCATTTC TGATGTGTTG     2220
ACCGCGAGTG ACTTTTGGGG CGGCGCCGGT TCGGCGGCCT GCCAGGGGTT CATTACCCAG     2280
CTGGGCCGTA ACTTCCAGGT GATNTACGAG CAGGCCAACG CCCACGGGCA GAAGGTGCAG     2340
GCTGCCGGCA ACAACATGGC ACAAACCGAC AGCGCCGTCG GCTCCAGCTG GGCATAAAGN     2400
TGGCTTAAGG CCCGCGCCGT CAATTACAAC GTGGCCGCAC ACCGGTTGGT GTGTGGCCAC     2460
GTTGTTATCT GAACGACTAA CTACTTCGAC CTGCTAAAGT CGGCGCGTTG ATCCCCGGTC     2520
GGATGGTGCT GAACTGGGAA GATGGCCTCA ATGCCCTTGT TGCGGAAGGG ATTGAGGCCA     2580
TCGTGTTTCG TACTTTAGGC GATCAGTGCT GGTTGTGGGA GTCGCTGCTG CCCGACGAGG     2640
TGCGCCGACT GCCCGAGGAA CTGGCCCGGG TGGACGCATT GTTGGACGAT CCGGCGTTCT     2700
TCGCCCCGTT CGTGCCGTTC TTCGACCCGC GCAGGGGCCG GCCGTCGACG CCGATGGAGG     2760
TCTATCTGCA GTTGATGTTT GTGAAGTTCC GCTACCGGCT GGCCTATGAG TCGCTGTGCC     2820
GGGAGGTGGC TGATTCGATC ACCTGACGGC GGTTTTGCCG CATTGCGCTG GACGGGTCGG     2880
TGCCGCATCC GACCACATTG ATGAAGCTCA CCACGCGTTG C                         2921

(2) INFORMATION FOR SEQ ID NO:8:

     (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 1704 base pairs
          (B) TYPE: nucleic acid
          (C) STRANDEDNESS: double
          (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

   (vi) ORIGINAL SOURCE:
          (A) ORGANISM: Mycobacterium tuberculosis

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

CGCGATCGTC GTCAACGARG TCGACCGTCA CCACGGACTG ATCAACAAGT TCGCAGGCGA       60
CGCCGCCCTG GCCATCTTCG GAGCCCCGAA CCGCCTCGAC CGTCCCGAAG ACGCCGCCT      120
GGCCCGCCGC CGGGCCATAN CCGANCGGCT GGCCNACGAG ATGCCCGAGG TCCAAGCCGG     180
CATCGRRGTG GCGGCAGGCC ANATCGTCGC CGGCAATGTC GGCGCCAAGC AAAGATTCNA     240
ATACACAGTG GTCGGCAAGC CGGTCAACCA NGCGGCCCGA TTGTGCGAAC TGGCCAAATC     300
ACACCCCGCG CGATTGGGTC TCGCCCGCTC GGCTCATGGT CACCCAATTC AAGGACTACT     360
TTGGCCTGGC GCACGACCTG CCGAAGTGGG CGAGTGAAGG CGCCAAAGCC GCCGGTGAGG     420
CCGCCAAGGC GTTGCCGGCC GCCGTTCCGG CCATTCCGAG TGCTGGCCTG AGCGGCGTTG     480
CGGGCGCCGT CGGTCAGGCG GCGTCGGTCG GGGGATTGAA GGTTCCGGCC GTTTCGACCG     540
CCACGACCCC GGCGGCGAGC CCCGCGGTGC TGGCGGCGTC CAACGGCCTC GGAGCCGCGG     600
CCGCCGCTGA AGGTTCGACA CACGCGTTTG GCGGGATGCC GCTCATGGGT ANCGGTGCCG     660
GACGTGCGTT TAACAACTTC GCTGCCCCTC GATACGGATT CAAGCCGACC GTGATCGCCC     720
AACCGCCGGC TGGCGGATGA CCAACTACGT TCGTTGATCG AGGATCGAAT TCNACGATTC     780
AAAGGGAGGA ATTCATATGA CCTCNCGTTT TATGACGGAT CCGCACGCNA TNCGGGACAT     840
GGCGGGGCGT TTTGAGGTGC ACGCCCAGAC GGTGGAGGAC GAGGCTNGCN GGATGTGGGC     900
GTCCGCGCAA AACATTTCCG GTGCGGGCTG GAGTGGCATG GCCGAGGCGA CCTCGNTAGA     960
CACCATGGCC CAGATGAATC AGGCGTTTCN CAACATCGTG AACATGCTGC ACGGGGTGNG    1020
TGACGGGCTG GTTCGCGACG CCAACAACTA CGAACAGCAA GAGCAGGCCT CCCAGCAGAT    1080
CCTCAGCAGC TGACCCGGCC CGACGACTCA GGAGGACACA TGACCATCAA CTATCAATTC    1140
GGGGACGTCG ACGCTCATGG CGCCATGATC CGCGCYNTGG CCGGGTTGCT GGAGGCCGAG    1200
CATCAGGCCA TCATTTCTGA TGTGTTGACC GCGAGTGACT TTTGGGGCGG CGCCGGTTCG    1260
GCGGCCTGCC AGGGGTTCAT TACCCAGTTG GGCCGTAACT TCCAGGTGAT TTACGAGCAG    1320
GCCAACGCCC ACGGGCAGAA GGTGCAGGCT GCCGCGCAACA ACATGGCACA AACCGACAGC    1380
GCCGTNGGNT CCAGCTGGGC CTAACCCGGG TCNTAAGTTG GGTCCGCGCA GGGCGGGCCG    1440
```

```
ATCAGCGTNG ACTTTGGCGC CCGATACACG GGCATNTTNT NGTCGGGAAC ACTGCGCCCG    1500
CGTCAGNTGC CCGCTTCCCC TTGTTNGGCG ACGTGCTCGG TGATGGCTTT GACGACCGCT    1560
TCGCCGGCGC GGCCAATCAA TTGGTCGCGC TTGCCTNTAG CCCATTCGTG CGACGCCCGC    1620
GGCGCCGCGA GTTGTCCCTT GAAATAAGGA ATCACAGCAC GGGCGAACAG CTCATAGGAG    1680
TGAAAGGTTG CCGTGGCGGG GCCC                                           1704

(2) INFORMATION FOR SEQ ID NO:9:

     (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 2286 base pairs
          (B) TYPE: nucleic acid
          (C) STRANDEDNESS: double
          (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA (genomic)

   (vi) ORIGINAL SOURCE:
          (A) ORGANISM: Mycobacterium tuberculosis

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CCGTCTTGGC GTCTGGGCGC ATTGTGATCT GGGCCANTTG CCCCTCCACC CAGACCGCGC      60
CCAGCTTGTC GATCCAGCCC GCGACCCGGA TTGCCACCGC GCGAACCGGG AACGGATTCT     120
CCGCTGAATT CTGGGTCACT TCGCAGTCGC GCGGGTGATC CTGTTGGCGA NCAGCGTCTG     180
GAACTGGCGT CNAACGCGTG CCGTAAGCCC AGCGTGTACG CCGTCAGCCC GACGCCGATG     240
CCGAATGCCT TGCCCGCCCAA GCTGAGCCGC GCCGGCTCCA CCAAGAGCGT CACGGTGAGC     300
CAGCCAACCA GATGCAAGGC GACGATCACC GCGAAGTGCC GAATTCGGCA CGAGAGGTGC     360
TGGAAATCCA GCAATACGCC CGCGAGCCGA TCTCGTTGGA CCAGACCATC GGCGACGANG     420
GCGACAGNCA GCTTGGCGAT TTCATCGAAA ACAGCGAGGC GGTGGTGGNC GTCGACGCGG     480
TGTCCTTCAC TTTGCTGCAT GATCAACTGC ANTCGGTGCT GGACACGCTC TCCGAGCGTG     540
AGGCGGGCGT GGTGCCGCTA CGCTTCGGCC TTACCGACGG CCAGCCGCGC ACCCTTGACG     600
AGATCGGCCA GGTCTACGGC GTGACCCGGG AACGCATCCG CCAGATCGAA TCCAAGACTA     660
TGTCGAAGTT GCGCCATCCG AGCCGCTCAC AGGTCCTGCG CGACTATCGT GCCGAATTCG     720
GCACGAGCCG TTTTGAGGTG CACGCCCAGA CGGTGGAGGA CGAGGCTCGC CGGATGTGGG     780
CGTCCGCGCA AAACATTTCC GGTGCGGGCT GGAGTGGCAT GGCCGANGCG ACCTCGCTAG     840
ACACCATGGC CCAGATGAAT CAGGCGTTTC GCAACATCGT GAACATGCTG CACGGGTTGC     900
GTGACCGGCT GGTTCGCGAC GCCNACAACT ACGAACAGCA AGAGCAGGCC TCCCAGCAGA     960
TCCTCAGCAG CTGACCCGGC CCGACGACTC AGGAGGACAC ATGACCATCA ACTATCAATT    1020
CGGGGACGTC GACGCTCATG GCGCCATGAT CCGCGCTCTG GCCGGGTTGC TGGAGGCCGA    1080
GCATCAGGCC ATCATTTCTG ATGTGTTGAC CGCGAGTGAC TTTTGGGGCG GCGCCGGTTC    1140
GGCGGCCTGC CAGGGGTTCA TTACCCAGTT GGGCCGTAAC TTCCAGGTGA TCTACGAGCA    1200
GGCCAACGCC CACGGGCAGA AGGTGCAGGC TGCCGGCAAC AACATGGCAC AAACCGACAG    1260
CGCCGTCGGC TCCAGCTGGG CCTAACCCGG GTCCTAAGTT GGGTCCGCGC AGGGCGGGCC    1320
GATCAGCGTC GACTTTGGCG CCCGATACAC GGGCATGTNG TNGTCGGGAA CACTGCGCCC    1380
GCGTCAGCTG CCCGCTTCCC CTTGTTCGGC GACGTGCTCG GTGATGGCTT TGACGACCGC    1440
TTCGCCGGCG CGGCCAATCA ATTGGTCGCG CTTGCCTCTA GCCTCGTGCC GAATTCGGCA    1500
CGAGGGTGCT GGTGCCGCGC TATCGGCAGC ACGTGAGCTC CACGACGAAC TCATCCCAGT    1560
GCTGGGTTCC GCGGAGTTCG GCATCGGCGT GTCGGCCGGA AGGGCCATCG CCGGCCACAT    1620
CGGCGCTCAA GCCGGCTTCG AGTACACCGT CATCGGCGAC CCGGTCAACG AGGCCGCCCG    1680
GCTCACCGAA CTGGCCAAAG TCGAGGATGG CCACGTTCTG GCGTCGGCGA TCGCGGTCAG    1740
TGGCGCCCTG GACGCCGAAG CATTGTGTTG GGATGTTGGC GAGGTGGTTG AGCTCCGCGG    1800
ACGTGCTGCA CCCACCCAAC TAGCCAGGCC AATGAATNTG GCNGCACCCG AAGAGGTTTC    1860
CAGCGAAGTA CGCGGCTAGT CGCGCTTGGC TGCNTTCTTC GCCGGCACCT TCCGGGCAGC    1920
TTTCCTGGCT GGCCGTTTTG CCGGACCCCG GGCTCGGCGA TCGGCCAACA GCTCGGCGGC    1980
GCGCTCGTCG GTTATGGAAG CCACGTNGTC GCCCTTACGC AGGCTGGCAT TGGTCTCACC    2040
```

GTCGGTGACG TACGGCCCGA ATCGGCCGTC CTTGATGACC ATTGGCTTGC CAGACGCCGG 2100
ATNTGNTCCC AGCTCGCGCA GCGGCGGAGC CGAAGCGCTT TGCCGGCCAC GACNTTTCGG 2160
CTCTGNGTAG ATNTTCAGGG CTTCGTCGAG CGNGATGGTG AATATATCGT CTTCGGTGAC 2220
CAGTGATCGA GAATCGTTGC CGCGCTTTAG ATACGGTCNG TAGCGCCCGT TCTGCGCGGT 2280
GATNTC 2286

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1136 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GGGCATCTTC CCCGACCGCG CCTCGATCAT CCGCCTCGTC GGAGCCGTCC TCGCCGAACA 60
ACACGACGAA TGGATCGAAG GACGGCGCTA CCTGGGCCTC GACGTCCTCA CCCGAGCCCG 120
AGCAGCACTG ACCAGCACCG AAGAACCGCC AAGCAGCAAA CCACCAACAC CCCAGCACTG 180
ACCACCTAGA CTGCCACCCG AAGGATCACG CGAGGAACCT TCACTCGTAC ACCACGTCCC 240
TNGCCTTGGC CTGGTNTCAG GCCCAGCTGG AGCCGACGGC GCTGTCGGTT TGCNCCATGT 300
TGTTGCCGGC AGCCTGCACC TTCTGCCCGT GGGCGTTGGC CTGCTCGTAG ATCACCTGGA 360
AGTTACGGCC CAACTGNGTA ATGAACCCCT GGCAGGCCGC CGAACCGGCG CCGCCCCAAA 420
AGTCACTCGC GGTCAACACA TCACGAATGA TGGCCTGATG CTCGGCCTCC AGCAACCCGG 480
CCTGAGCGCG GATCATGGCG CCGTGAGCGT CGACATCACC GAACTGATAG TTGATGGTCA 540
TCGAACCTGT TCTCCTTCGC TTGTAAAAGT ATTGTGCTGC AGCGGCTGAC GTTAGCTGCT 600
GAGGATCTGC TGGGAGGCCT GCTCTTGCCT CGTGCCGAAT TCGGCACGAG AGGCCGCCTT 660
CGAACGAAATC CTTTGAGAAT TCGCCAAGGC CGTCGACCCA GCATGGGGTC AGCTCGCCAG 720
CCGCGCCGGC TGGCAACCGT TCCCGCTCGA GAAAGACCTG GAGGAATACC AGTGACAAAC 780
GACCTCCCAG ACGTCCGAGA GCGTGACGGC GGTCCACGTC CCGCTCCTCC TGCTGGCGGG 840
CCACGCTTGT CAGACGTGTG GGTTTACAAC GGGCGGCGCT ACGACCTCAG TGAGTGGATT 900
TCCAAGCATC CCGGCGGCGC CTTNTTCATT GGGCGGACCA AGAACCGCGA CATCACCGCA 960
ATCGTCAAGT CCTACCATCG TGATCCGGCG ATTGTCGAGC GAATCCTGCA GCGGAGGTAC 1020
GCGTTGGGCC GCGACGCAAC CCCTAGGGAC ATCCACCCCA AGCACAATGC ACCGGCATTT 1080
CTGTTCAAAG ACGACTTCAA CAGCTGGCGG GACACCCCGA AGTATCGATT NGACGA 1136

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 967 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```
TGAGCGCCAA CCCTACCGTC GGTTCGTCAC ACGGACCGCA TGGCCTGCTC CGCGGACTGC      60
CGCTAGGGTC GCCGATCACT CGGCGTAGCG GCGCCTTTGC CCACCGATAT GGGTTCCGTC     120
ACAGTGTGGT TGCCCCGCCG CCATCGGCCG GATAACGCCA TGACCTCAGC TCGGCAGAAA     180
TGACAATGCT CCCAAAGGCG TGAGCACCCG AAGACAACTA AGCAGGAGAT CGCATGCCGT     240
TTGTGACTAC CCAACCAGAA GCACTGGCGG CGGCGGCCGG CAGTCTGCAG GGAATCGGCT     300
CCGCATTGAA CGCCCAGAAT GCGGCTGCGG CGACTCCCAC GACGGGGGTG GTCCGGCGGC     360
CGCCGATGAA NTGTCGGCNC TGACGGCGGC TCAGTTCGCG GCACACGCCC AGATCTATCA     420
GGCCGTCAGC GCCCAGGCCG CGGCGATTCA CGAGATGTTC GTCAACACTC TACAGATGAG     480
CTCAGGGTCG IATGCTGCTA CCGAGGCCGC CAACGCGGCC GCGGCCGGNT AGAGGAGTCA     540
CTGCGATGGA TTTTGGGGCG TTGCCGCCGG AGGTCAATTC GGTGCGGATG IATGCCGTTC     600
CTGGCTCGGC ACCAATGGTC GCTGCGGCGT CGGCCTGGAA CGGGTTGGCC GCGGAGCTGA     660
GTTCGGCCGC CACCGCTTAT GAGACGGTGA TCACTCAGCT CAGCAGTGAG GGGTGGCTAG     720
GTCCGGCGTC AGCGGCGAIG GCCGAGGCAG TTGCGCCGTA TGTGGCGTGG ATGAGTGCCG     780
CTGCGGCGCA AGCCGAGCAG GCGGCCACAC AGGCCAGGGC CGCCGCGGCC GCTTTTGAGG     840
CGGCGTTTGC CGCGACGGTG CCTCCGCCGT TGATCGCGGC CAACCGGGCT TCGTTGATGC     900
AGCTGATCTC GACGAATGTC TTTGGTCAGA ACACCTCGGC GATCGCGGCC GCCGAAGCTC     960
AGTACGG                                                              967
```

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 585 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

```
TGGATTCCGA TAGCGGTTTC GGCCCCTCGA CGGGCGACCA CGGCGCGCAG GCCTCCGAAC      60
GGGGGCCCGG GACGCTGGGA TTCGCCNGGA CCGCAACCAA AGAACGCCGG GTCCGGGCGG     120
TCGGGCTGAC CGCACTGGCC GGTGATGAGT TCGGCGACGG CCCCCCGATG CCGATGGTGC     180
CGGGGACCTG GGAGCAGGGC AGCAACGAGC CCGAGGCGCC CGACGGATCG GGGACGAGGG     240
GAGGCGACGG CTTACCGCAC GACAGCAAGT AACCGBATTC CGAATCACGT GGACCCGTAC     300
GGGTCGAAAG GACAGATGTT ATGACCCTTT TGGATGCTCA TATCCCACAG TTGGTGGCCT     360
CCCAGTCGGC GTTTGCCGCC AAGGCGGGGC TGATGCGGCA CACGATCGGT CAGGCCGAGC     420
AGGCGGCGAT GTCGGCTCAG GCGTTTCACC AGGGGGAGTC GTCGGCGGCG TTTCAGGCCG     480
CCCATGCCCG GTTTGTGGCG GCGGCCGCCA AAGTCAACAC CTTGTTGGAT GTCGCGCAGG     540
CGAATCTGGG TGAGGCCGCC GGTACCTAIG TGGCCGCCGA TGCTG                     585
```

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 144 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Ala Leu Val Thr Thr Asn Phe Phe Gly Val Asn Thr Ile Pro Ile Ala
1 5 10 15
Leu Asn Glu Ala Asp Tyr Leu Arg Met Trp Ile Gln Ala Ala Thr Val
20 25 30
Met Ser His Tyr Gln Ala Val Ala His Glu Ile Trp Cys Leu His Glu
35 40 45
Xaa Ala Ser Ser Gly Lys Pro Trp Ala Ser Ile Thr Thr Gly Ala Pro
50 55 60
Gly Ser Pro Ala Ser Thr Thr Arg Ser Arg Thr Pro Leu Val Ser Thr
65 70 75 80
Asn Arg Xaa Val Xaa Ala Pro Ile Val Ser Pro Asn His Thr Gly His
85 90 95
Arg Pro Glu Lys Gly Leu Gly Ser Xaa Gln Arg Arg Leu Ser Arg Val
100 105 110
Leu Pro Arg Ile Ile Asp Arg Pro Ala Gly Pro Xaa Gly Pro Pro Leu
115 120 125
Thr Ser Gly Ser His Phe Leu Cys Ser Trp His Gly Tyr Ser Ser Gln
130 135 140

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 352 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

His Ala Leu Ala Ala Gln Tyr Thr Glu Ile Ala Thr Glu Leu Ala Ser
1 5 10 15
Val Leu Ala Ala Val Gln Ala Ser Ser Trp Gln Gly Pro Ser Ala Asp
20 25 30
Arg Phe Val Val Ala His Gln Pro Phe Arg Tyr Trp Leu Thr His Ala
35 40 45
Ala Thr Val Ala Thr Ala Ala Ala Ala Ala His Xaa Thr Ala Ala Ala
50 55 60
Gly Tyr Thr Ser Ala Leu Gly Gly Met Pro Thr Leu Ala Glu Leu Ala
65 70 75 80
Ala Asn His Ala Met His Gly Ala Leu Val Thr Thr Asn Phe Phe Gly
85 90 95
Val Asn Thr Ile Pro Ile Ala Leu Asn Glu Ala Asp Tyr Leu Arg Met
100 105 110
Trp Ile Gln Ala Ala Thr Val Met Ser His Tyr Gln Ala Val Ala His
115 120 125

```
Glu Ser Val Ala Ala Thr Pro Ser Thr Pro Pro Ala Pro Gln Ile Val
    130                 135                 140
Thr Ser Ala Ala Ser Ser Ala Ala Ser Ser Ser Phe Pro Asp Pro Thr
145                 150                 155                 160
Lys Leu Ile Leu Gln Leu Leu Lys Asp Phe Leu Glu Leu Leu Arg Tyr
                165                 170                 175
Leu Ala Val Glu Leu Leu Pro Gly Pro Leu Gly Asp Leu Ile Ala Gln
            180                 185                 190
Val Leu Asp Trp Phe Ile Ser Phe Val Ser Gly Pro Val Phe Thr Phe
        195                 200                 205
Leu Ala Tyr Leu Val Leu Asp Pro Leu Ile Tyr Phe Gly Pro Phe Ala
    210                 215                 220
Pro Leu Thr Ser Pro Val Leu Leu Pro Ala Val Glu Leu Arg Asn Arg
225                 230                 235                 240
Leu Lys Thr Ala Thr Gly Leu Thr Leu Pro Pro Thr Val Ile Phe Asp
                245                 250                 255
His Pro Thr Pro Thr Ala Val Ala Glu Tyr Val Ala Gln Gln Met Ser
            260                 265                 270
Gly Ser Arg Pro Thr Glu Ser Gly Asp Pro Thr Ser Gln Val Val Glu
        275                 280                 285
Pro Ala Arg Ala Glu Phe Gly Thr Ser Ala Val His Gln Ile Pro Pro
    290                 295                 300
Arg Pro Ala Asp Thr Arg Arg Ala Cys Arg His Arg Asp Asp Val Pro
305                 310                 315                 320
Arg Asp Ser Arg Ile Ala Gln His Arg Asp Gly Ala Gly Leu Asp Pro
                325                 330                 335
Thr Glu Arg Gly Thr Ser Glu Gly Asp Gln Gly Leu Val Ser Gly Trp
            340                 345                 350
```

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 141 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

```
Met Asp Phe Gly Ala Leu Pro Pro Glu Val Asn Ser Val Arg Met Tyr
1               5                   10                  15
Ala Val Pro Gly Ser Ala Pro Met Val Ala Ala Ala Ser Ala Trp Asn
            20                  25                  30
Gly Leu Ala Ala Glu Leu Ser Ser Ala Ala Thr Gly Tyr Glu Thr Val
        35                  40                  45
Ile Thr Gln Leu Ser Ser Glu Gly Trp Leu Gly Pro Ala Ser Ala Ala
    50                  55                  60
Met Ala Glu Ala Val Ala Pro Tyr Val Ala Trp Met Ser Ala Ala Ala
65                  70                  75                  80
Ala Gln Ala Glu Gln Ala Ala Thr Gln Ala Arg Ala Ala Ala Ala Ala
```

```
                85                  90                  95
Phe Glu Ala Ala Phe Ala Ala Thr Val Pro Pro Pro Leu Ile Ala Ala
            100                 105                 110
Asn Arg Ala Ser Leu Met Gln Leu Ile Ser Thr Asn Val Phe Gly Gln
        115                 120                 125
Asn Thr Ser Ala Ile Ala Ala Ala Glu Ala Gln Tyr Gly
    130                 135                 140
```

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 58 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

```
Met Ala Ser Arg Phe Met Thr Asp Pro His Ala Met Arg Asp Met Ala
1               5                  10                  15
Gly Arg Phe Glu Val His Ala Gln Thr Val Glu Asp Glu Ala Arg Arg
            20                  25                  30
Met Trp Ala Ser Ala Gln Asn Ile Ser Gly Ala Gly Trp Ser Gly Met
        35                  40                  45
Ala Glu Ala Thr Ser Leu Asp Thr Met Thr
    50                  55
```

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 67 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

```
Met Thr Ile Asn Tyr Gln Phe Gly Asp Val Asp Ala His Gly Ala Met
1               5                  10                  15
Ile Arg Ala Gln Ala Ala Ser Leu Glu Ala Glu His Gln Ala Ile Val
            20                  25                  30
Arg Asp Val Leu Ala Ala Gly Asp Phe Trp Gly Gly Ala Gly Ser Val
        35                  40                  45
Ala Cys Gln Glu Phe Ile Thr Gln Leu Gly Arg Asn Phe Gln Val Ile
    50                  55                  60
```

Tyr Glu Gln
65

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 58 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Met Ala Ser Arg Phe Met Thr Asp Pro His Ala Met Arg Asp Met Ala
1 5 10 15
Gly Arg Phe Glu Val His Ala Gln Thr Val Glu Asp Glu Ala Arg Arg
20 25 30
Met Trp Ala Ser Ala Gln Asn Ile Ser Gly Ala Gly Trp Ser Gly Met
35 40 45
Ala Glu Ala Thr Ser Leu Asp Thr Met Thr
50 55

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 94 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Met Thr Ile Asn Tyr Gln Phe Gly Asp Val Asp Ala His Gly Ala Met
1 5 10 15
Ile Arg Ala Gln Ala Ala Ser Leu Glu Ala Glu His Gln Ala Ile Val
20 25 30
Arg Asp Val Leu Ala Ala Gly Asp Phe Trp Gly Gly Ala Gly Ser Val
35 40 45
Ala Cys Gln Glu Phe Ile Thr Gln Leu Gly Arg Asn Phe Gln Val Ile
50 55 60
Tyr Glu Gln Ala Asn Ala His Gly Gln Lys Val Gln Ala Ala Gly Asn
65 70 75 80
Asn Met Ala Gln Thr Asp Ser Ala Val Gly Ser Ser Trp Ala
85 90

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Asn Met Leu His Gly Val Arg Asp Gly Leu Val Arg Asp Ala Asn Asn
1 5 10 15
Tyr Glu Gln Gln Glu Gln Ala Ser Gln Gln Ile Leu Ser Ser
20 25 30

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 94 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

Met Thr Ile Asn Tyr Gln Phe Gly Asp Val Asp Ala His Gly Ala Met
1 5 10 15
Ile Arg Ala Gln Ala Gly Leu Leu Glu Ala Glu His Gln Ala Ile Ile
20 25 30
Arg Asp Val Leu Thr Ala Ser Asp Phe Trp Gly Gly Ala Gly Ser Ala
35 40 45
Ala Cys Gln Gly Phe Ile Thr Gln Leu Gly Arg Asn Phe Gln Val Ile
50 55 60
Tyr Glu Gln Ala Asn Ala His Gly Gln Lys Val Gln Ala Ala Gly Asn
65 70 75 80
Asn Met Ala Gln Thr Asp Ser Ala Val Gly Ser Ser Trp Ala
85 90

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 69 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

```
Ala Arg Arg Met Trp Ala Ser Ala Gln Asn Ile Ser Gly Ala Gly Trp
1               5                  10                  15
Ser Gly Met Ala Glu Ala Thr Ser Leu Asp Thr Met Ala Gln Met Asn
            20                  25                  30
Gln Ala Phe Arg Asn Ile Val Asn Met Leu His Gly Val Arg Asp Gly
        35                  40                  45
Leu Val Arg Asp Ala Asn Asn Tyr Glu Gln Gln Glu Gln Ala Ser Gln
    50                  55                  60
Gln Ile Leu Ser Ser
65
```

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 94 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

```
Met Thr Ile Asn Tyr Gln Phe Gly Asp Val Asp Ala His Gly Ala Met
1               5                  10                  15
Ile Arg Ala Gln Ala Gly Leu Leu Glu Ala Glu His Gln Ala Ile Ile
            20                  25                  30
Arg Asp Val Leu Thr Ala Ser Asp Phe Trp Gly Gly Ala Gly Ser Ala
        35                  40                  45
Ala Cys Gln Gly Phe Ile Thr Gln Leu Gly Arg Asn Phe Gln Val Ile
    50                  55                  60
Tyr Glu Gln Ala Asn Thr His Gly Gln Lys Val Gln Ala Ala Gly Asn
65                  70                  75                  80
Asn Met Ala Gln Thr Asp Ser Ala Val Xaa Ser Ser Trp Ala
                85                  90
```

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 52 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

```
Gly Met Ala Glu Ala Thr Ser Xaa Asp Thr Met Thr Gln Met Asn Gln
1               5                  10                  15
Ala Phe Arg Asn Ile Val Asn Met Leu His Gly Val Arg Asp Gly Leu
            20                  25                  30
Val Arg Asp Ala Asn Xaa Tyr Glu Gln Gln Glu Gln Ala Ser Gln Gln
        35                  40                  45
Ile Leu Ser Ser
    50
```

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 94 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

```
Met Thr Ile Asn Tyr Gln Phe Gly Asp Val Asp Ala His Gly Ala Met
1               5                  10                  15
Ile Arg Ala Gln Ala Gly Ser Leu Glu Ala Glu His Gln Ala Ile Ile
            20                  25                  30
Ser Asp Val Leu Thr Ala Ser Asp Phe Trp Gly Gly Ala Gly Ser Ala
        35                  40                  45
Ala Cys Gln Gly Phe Ile Thr Gln Leu Gly Arg Asn Phe Gln Val Xaa
    50                  55                  60
Tyr Glu Gln Ala Asn Ala His Gly Gln Lys Val Gln Ala Ala Gly Asn
65                  70                  75                  80
Asn Met Ala Gln Thr Asp Ser Ala Val Gly Ser Ser Trp Ala
                85                  90
```

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 98 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

```
Met Thr Ser Arg Phe Met Thr Asp Pro His Ala Met Arg Asp Met Ala
1               5                  10                  15
Gly Arg Phe Glu Val His Ala Gln Thr Val Glu Asp Glu Ala Arg Arg
            20                  25                  30
Met Trp Ala Ser Ala Gln Asn Ile Ser Gly Ala Gly Trp Ser Gly Met
        35                  40                  45
Ala Glu Ala Thr Ser Leu Asp Thr Met Ala Gln Met Asn Gln Ala Phe
    50                  55                  60
Arg Asn Ile Val Asn Met Leu His Gly Val Arg Asp Gly Leu Val Arg
65                  70                  75                  80
Asp Ala Asn Asn Tyr Glu Gln Gln Glu Gln Ala Ser Gln Gln Ile Leu
                85                  90                  95
Ser Ser
```

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 94 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

```
Met Thr Ile Asn Tyr Gln Phe Gly Asp Val Asp Ala His Gly Ala Met
1               5                  10                  15
Ile Arg Ala Xaa Ala Gly Leu Leu Glu Ala Glu His Gln Ala Ile Ile
            20                  25                  30
Ser Asp Val Leu Thr Ala Ser Asp Phe Trp Gly Gly Ala Gly Ser Ala
        35                  40                  45
Ala Cys Gln Gly Phe Ile Thr Gln Leu Gly Arg Asn Phe Gln Val Ile
    50                  55                  60
Tyr Glu Gln Ala Asn Ala His Gly Gln Lys Val Gln Ala Ala Gly Asn
65                  70                  75                  80
Asn Met Ala Gln Thr Asp Ser Ala Val Gly Ser Ser Trp Ala
                85                  90
```

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 81 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

```
Arg Phe Glu Val His Ala Gln Thr Val Glu Asp Glu Ala Arg Arg Met
1               5                  10                  15
Trp Ala Ser Ala Gln Asn Ile Ser Gly Ala Gly Trp Ser Gly Met Ala
            20                  25                  30
Xaa Ala Thr Ser Leu Asp Thr Met Ala Gln Met Asn Gln Ala Phe Arg
        35                  40                  45
Asn Ile Val Asn Met Leu His Gly Val Arg Asp Gly Leu Val Arg Asp
    50                  55                  60
Ala Asn Asn Tyr Glu Gln Gln Glu Gln Ala Ser Gln Gln Ile Leu Ser
65                  70                  75                  80
Ser
```

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 94 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

```
Met Thr Ile Asn Tyr Gln Phe Gly Asp Val Asp Ala His Gly Ala Met
1               5                  10                  15
Ile Arg Ala Leu Ala Gly Leu Leu Glu Ala Glu His Gln Ala Ile Ile
            20                  25                  30
Ser Asp Val Leu Thr Ala Ser Asp Phe Trp Gly Gly Ala Gly Ser Ala
        35                  40                  45
Ala Cys Gln Gly Phe Ile Thr Gln Leu Gly Arg Asn Phe Gln Val Ile
    50                  55                  60
Tyr Glu Gln Ala Asn Ala His Gly Gln Lys Val Gln Ala Ala Gly Asn
65                  70                  75                  80
Asn Met Ala Gln Thr Asp Ser Ala Val Gly Ser Ser Trp Ala
                85                  90
```

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

Gln Glu Gln Ala Ser Gln Gln Ile Leu Ser Ser
1 5 10

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 94 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

Met Thr Ile Asn Tyr Gln Phe Gly Asp Val Asp Ala His Gly Ala Met
1 5 10 15
Ile Arg Ala Gln Ala Gly Leu Leu Glu Ala Glu His Gln Ala Ile Ile
20 25 30
Arg Asp Val Leu Thr Ala Ser Asp Phe Trp Gly Gly Ala Gly Ser Ala
35 40 45
Ala Cys Gln Gly Phe Ile Thr Gln Leu Gly Arg Asn Phe Gln Val Ile
50 55 60
Tyr Glu Gln Ala Asn Ala His Gly Gln Lys Val Gln Ala Ala Gly Asn
65 70 75 80
Asn Met Ala Gln Thr Asp Ser Ala Val Gly Ser Ser Trp Ala
85 90

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 99 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mycobacterium tuberculosis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

Met Ser Phe Val Thr Thr Gln Pro Glu Ala Leu Ala Ala Ala Ala Ala
1 5 10 15
Asn Leu Gln Gly Ile Gly Thr Thr Met Asn Ala Gln Asn Ala Ala Ala